June 1967

ON PARTITIONING A SET OF NORMAL POPULATIONS BY THEIR LOCATIONS WITH RESPECT TO A CONTROL USING A SINGLE-STAGE, A TWO-STAGE AND A SEQUENTIAL PROCEDURE*

by

Yung Liang Tong

Technical Report No. 90

University of Minnesota
Minneapolis, Minnesota

* Research supported by National Science Foundation Grant No. NSF-GP-3813.

Acknowledgement

The author wishes to express his deep gratitude to Professor Milton Sobel for providing guidance and continuing encouragement during the course of this research. The author also wishes to thank Professor Norman Starr, Carnegie Institute of Technology, for a helpful discussion.

TABLE OF CONTENTS

CHAPTER I Introduction and Summary

1.1 A Short History of the Problem.

During the last two decades a considerable amount of research has been devoted to the area of ranking and selection problems. It represents an alternative to the classical theory of testing statistical hypothesis and the philosophy of this approach was discussed by Bechhofer [3].

At the early stage of the development, most of the problems considered were concerned with the selection of a subset containing the populations associated with the largest (or smallest) population parameters such that the size of the subset was a predetermined constant and the parametric form of the populations did not involve any nuisance parameter. Bechhofer [3] in 1954 proposed a decision rule to select the "best" population (the population associated with the largest population mean) from a set of k normal populations with a common known variance. In the same year, Bechhofer and Sobel [6] considered the problem of selecting the population associated with the smallest variance from k normal populations. Other authors had worked on this type of problem for different parametric forms. For a general survey of the literatures, see Eaton [12].

It appears that Paulson [23] was the first author to consider the problem of comparing a set of populations with a standard or control. His main interest was to select the best population which could be either the control or some other. In 1955 Dunnett [10] treated the problem of deriving a joint confidence region for the differences between the control population mean and each of the other population means from a set of normal populations with a common unknown variance. Both Paulson and Dunnett considered a particular slippage configuration only and satisfied a probability requirement in the null case when all the populations are identical. Based on a subset formulation, Gupta and Sobel [17] proposed a rule to select a subset which contains all populations better than a standard. One of the significant differences between their problem and the previous ones is that the size of the subset selected is a chance variable, and the evaluation of the proposed rule is mainly based on the expected subset size. The same problem and its applicability was also discussed by Lehmann [19] from the indifference zone approach within the general framework of model I problems of selection.

1.2 The Problem, the Approaches and Main Results.

In this work we consider the problem of partitioning a set of normal populations into two subsets according to their locations with respect to a control population, based on indifference zone approach. Let $\pi_0, \pi_1, \ldots, \pi_k$ be $(k + 1)$ normal populations with means $\mu_0, \mu_1, \ldots \mu_k$ and a common variance σ^2; and let π_0 denote the standard or control population. In an experiment the population π_0 could be either a dummy population which represents a certain fixed standard or a treatment whose response is also under investigation. Hence the true value of μ_0 could be either known or unknown. For arbitrary but fixed constants δ_1^* and δ_2^* such that $\delta_1^* \leq \delta_2^*$, we define three disjoint and exhaustive subsets Ω_W, Ω_I and Ω_B of the set

$$\Omega = \{\pi_1, \pi_2, \ldots, \pi_k\} \tag{1.1}$$

by

$$\begin{aligned} \Omega_W &= \{\pi_i : \mu_i \leq \mu_0 + \delta_1^*\} \\ \Omega_I &= \{\pi_i : \mu_0 + \delta_1^* < \mu_i < \mu_0 + \delta_2^*\} \\ \Omega_B &= \{\pi_i : \mu_i \geq \mu_0 + \delta_2^*\}. \end{aligned} \tag{1.2}$$

After observations have been taken, the set Ω is partitioned into two disjoint subsets S_W and S_B. We define a correct decision by

Definition 1.1.

A decision is a correct decision (CD) if $\Omega_W \subset S_W$ and $\Omega_B \subset S_B$.

An equivalent definition to Definition 1.1 is that $S_W \subset (\Omega_W \cup \Omega_I)$

and $S_B \subset (\Omega_B \cup \Omega_I)$. It is noted that the open interval $(\mu_0 + \delta_1^*, \mu_0 + \delta_2^*)$ is considered as the indifference zone and a correct decision puts no restrictions on those π_i's in the set Ω_I; i.e., they can be put either in S_W or S_B without coutributing any positive loss. With this consideration, it will be consistent to give the following

Definition 1.2.

A population $\pi_i \in \Omega$ is misclassified if $\pi_i \in (\Omega_W \cap S_B) \cup (\Omega_B \cap S_W)$.

The practical implication of this problem is easily seen. In many cases the experimenter desires to partition k different treatments into two disjoint subsets: one is worse than the control and the other better. This can be achieved by setting $\delta_1^* < 0$ and $\delta_2^* > 0$. For some other applications, he may want to partition the populations into one subset whose population means are greater than μ_0 by at most δ_1^*, and another subset whose population means are greater than μ_0 by at least δ_2^* (where $\delta_1^* < \delta_2^*$). In this case we have both $\delta_1^* > 0$ and $\delta_2^* > 0$. The corresponding case in which both δ_1^* and δ_2^* are negative can be considered similarly.

Let P^* be an arbitrary but preassigned constant such that $2^{-k} < P^* \leq 1$. The statistical problem is to find a rule R which consists of a sampling procedure and a terminal decision rule such that the appropriate probability requirement below is satisfied.

(1) When σ^2 is known, R is such that

(1.3) $$P[CD|\vec{\mu}, \sigma^2; R] \geq P^* \quad \text{for every mean vector } \vec{\mu}.$$

(2) When σ^2 is unknown, R is such that

(1.4) $$P[CD|\vec{\mu}, \sigma^2; R] \geq P^* \quad \text{for every } \vec{\mu} \text{ and every } \sigma^2 > 0.$$

The case of known σ^2 is considered in Chapter II. A single-stage procedure analogous to that of Bechhofer [3] is used, and the proposed decision rule is proved to be Bayesian, minimax and admissible by applying Lehmann's theorem on multiple decision problems [18]. By the non-existence theorem of Dantzig [9], there is no single-stage procedure that can solve this problem when σ^2 is unknown. A two-stage procedure analogous to that originally proposed by Stein [30] and introduced to the area of ranking and selection problems by Bechhofer, Dunnett and Sobel [4] is used in Chapter III. The relative inefficiency of Stein's procedure has been discussed by a number of authors. It has been explained as partly due to the fact that the information of the observations in the second stage is not fully utilized in estimating the unknown σ^2. This gives some thought of performing the experiment so that the unknown σ^2 can be estimated sequentially. A random stopping rule has been developed to solve the problem considered by Stein, and its relative efficiency, especially when σ^2 is large, has been studied lately (see Anscombe [1], [2], Chow and Robbins [7], Gleser, Robbins and Starr [14] and Starr [29]). A procedure for the present problem based on this idea is developed in Chapter IV to serve as an alternative to the two-stage procedure.

The determination of the sample sizes required under each of the above procedures is also investigated. When the mean of the control population is unknown, the equi-coordinate percentage points of a multivariate normal distribution with mean vector 0 and a certain covariance (or correlation) matrix Σ defined in (2.20) and the equi-coordinate percentage points of a multivariate t distribution with the same correlation matrix Σ are needed for this purpose. Tables of these percentage points have been constructed which can be found at the end of this work.

In Appendix A a general result about certain multivariate normal distributions is proved; and this result is used in Chapter II to find the infimum of the probability of a correct decision under the proposed decision rule. The behavior of the percentage points of the multivariate t distributions with large degrees of freedom is examined in Section 3.3. For this purpose, a result about the relationship between convergence in distribution and convergence of the sequence of the corresponding percentage points for a fixed percentile is needed. This result is proved more generally in Appendix B.

The sample size required , the relative efficiency, the expected misclassification size and their asymptotic behavior for the single-stage, two-stage and sequential procedures are investigated and are shown to be functions of the percentage points of the above-mentioned multivariate normal and multivariate t distributions. It is interesting but not suprising to observe that when σ^2 is large, the sequential procedure is asymptotically relatively efficient and it has the same expected misclassification size as the single-stage procedure.

1.3 Assumptions and Notations.

The following assumptions are made throughout this work:

(1) there is no a priori knowledge regarding the means of the populations in the set Ω;

(2) the observations are taken a vector at a time; and

(3) the observations are independent.

Unless mentioned otherwise, the following notations will be adopted:

(1.5) $$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad -\infty < z < \infty;$$

(1.6) $$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx, \quad -\infty < z < \infty;$$

(1.7) $$m = \begin{cases} k/2 & \text{if } k \text{ is even,} \\ (k+1)/2 & \text{if } k \text{ is odd;} \end{cases}$$

(1.8) $$d = \frac{1}{2}(\delta_1^* + \delta_2^*);$$

(1.9) $$a = \frac{1}{2}(\delta_2^* - \delta_1^*); \text{ and}$$

(1.10) $$\lambda = \frac{\sigma}{a}.$$

CHAPTER II A Single-Stage Procedure

We have $(k + 1)$ normal populations $\pi_0, \pi_1, \ldots, \pi_k$ with means $\mu_0, \mu_1, \ldots, \mu_k$ and a common variance σ^2. In this chapter we assume that the true value of σ^2 is known. The statistical problem concerned is to find a sampling procedure and a terminal decision rule such that the probability requirement (1.3) is satisfied. It should be emphasized that we want the inequality in (1.3) to hold under all possible configurations of the mean vector $\vec{\mu}$. A single-stage procedure analogous to that of Bechhofer [3] is given as a solution to this problem. A rule is proposed and the probability of a correct decision (PCD) and its infimum are examined in Section 2.1. The sample size required to achieve the probability requirement under this sampling procedure is then determined. When μ_0 is unknown, the determination of the sample size depends on the equi-coordinate percentage points of a certain multivariate normal distribution (see (2.22)). These are tabulated at several probability levels in Table 1. In Section 2.2, an upper bound on this sample size is found based on Boole's inequality in terms of the percentage points of equicorrelated multivariate normal distributions. Lehmann [18] has developed a general theory for multiple decision problems; the present problem is fitted within this framework in Section 2.3 and it is proved there that the proposed decision rule has several optimal properties. The expected misclassification size and its supremum under the proposed rule is also investigated in Section 2.3.

2.1 The Proposed Rule and its PCD.

Since the observations are obtained a vector at a time, the distribution of the observations can be written in their multivariate form. For known μ_0 let

(2.1) $$\{X_{1j}, X_{2j}, \ldots, X_{kj}\}_{j=1}^{\infty}$$

be a sequence of independent vector observations from the population with joint density

(2.2) $$f(x_1, x_2, \ldots, x_k; \mu_1, \mu_2, \ldots, \mu_k, \sigma^2) = \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x_i - \mu_i)^2}$$

for $-\infty < x_i < \infty$ and with parameter spaces $-\infty < \mu_i < \infty$ $(i = 1, 2, \ldots, k)$.
For unknown μ_0, let

$$\text{(2.3)} \qquad \{X_{oj}, X_{1j}, \ldots, X_{kj}\}_{j=1}^{\infty}$$

be a sequence of independent vector observations from the population with joint density

$$\text{(2.4)} \qquad f(x_0, x_1, \ldots, x_k; \mu_0, \mu_1, \ldots, \mu_k, \sigma^2) = \prod_{i=0}^{k} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x_i - \mu_i)^2}$$

for $-\infty < x_i < \infty$ and with parameter spaces $-\infty < \mu_i < \infty$ $(i = 0, 1, \ldots, k)$.
Throughout this chapter, we assume that σ^2 is a known constant.

Definition 2.1. (1) Rule R_{1A}: when μ_0 is known, observe the sequence (2.1) for $j = 1, 2, \ldots, N$ where N is to be determined below. Compute $\bar{X}_i = \frac{1}{N}\sum_{j=1}^{N} X_{ij}$ for $i = 1, 2, \ldots, k$, and use the decision rule:

$$\text{(2.5)} \qquad \begin{aligned} S_W &= \{\pi_i | \pi_i \in \Omega, \bar{X}_i - \mu_0 < d\}, \\ S_B &= \{\pi_i | \pi_i \in \Omega, X_i - \mu_0 > d\}. \end{aligned}$$

(2) Rule R_{1B}: when μ_0 is unknown, observe the sequence (2.3) for $j = 1, 2, \ldots, N$ where N is to be determined below. Compute $\bar{X}_i = \frac{1}{N}\sum_{j=1}^{N} X_{ij}$ for $i = 0, 1, \ldots, k$, and use the decision rule:

$$\text{(2.6)} \qquad \begin{aligned} S_W &= \{\pi_i | \pi_i \in \Omega, \bar{X}_i - \bar{X}_0 < d\}, \\ S_B &= \{\pi_i | \pi_i \in \Omega, \bar{X}_i - \bar{X}_0 > d\}. \end{aligned}$$

Remark 2.1. Since for every $i = 1, 2, \ldots, k$, the event $[\bar{X}_i - \mu_0 = d]$ or $[\bar{X}_i - \bar{X}_0 = d]$ has probability zero, the equality is ignored in the above decision rule.

Once the sample size N is determined so that the probability requirement (1.3) is satisfied, then the rule specified in Definition 2.1 is completely defined. For this purpose we give the following

Definition 2.2. A mean vector $\vec{\mu}^0 = (\mu_0^0, \mu_1^0, \ldots, \mu_k^0)$ is a least favorable (LF) configuration under a rule R if

$$\text{(2.7)} \qquad P[CD|\vec{\mu}^0, \sigma^2; R] = \inf_{\vec{\mu}} P[CD|\vec{\mu}, \sigma^2; R].$$

<u>Lemma 2.1</u>. For Φ defined in (1.6),

$$\text{(2.8)} \qquad (1) \quad \inf_{\vec{\mu}} P[CD|\vec{\mu}, \sigma^2; R_{1A}] = \Phi^k(\sqrt{N}/\lambda);$$

(2) the LF configuration $\vec{\mu}^0$ under R_{1A} is such that either $\mu_i^0 = \mu_0 + \delta_1^*$ or $\mu_i^0 = \mu_0 + \delta_2^*$ $(i = 1,2,\ldots, k)$.

<u>Remark 2.2</u>. It should be noted that the LF configuration under R_{1A} is not unique in the sense that it does not depend on the number of μ_i^0's which are equal to $\mu_0 + \delta_1^*$.

<u>Proof of the lemma</u>: Let there be r populations in Ω_W, $(k - r - s)$ populations in Ω_I and s populations in Ω_B. Without loss of generality we can assume

$$\text{(2.9)} \qquad \begin{aligned} \Omega_W &= \{\pi_1, \pi_2,\ldots, \pi_r\}, \\ \Omega_I &= \{\pi_{r+1}, \pi_{r+2},\ldots, \pi_{k-s}\}, \text{ and} \\ \Omega_B &= \{\pi_{k-s+1}, \pi_{k-s+2},\ldots, \pi_k\}. \end{aligned}$$

Then, by Definitions 1.1,

$$P[CD|\vec{\mu}, \sigma^2; R_{1A}] = P[\bar{X}_i - \mu_0 < d, \bar{X}_j - \mu_0 > d \, (\substack{i=1,\, 2,\ldots,\, r \\ j=k-s+1,\, k-s+2,\ldots,k})|\vec{\mu}, \sigma^2]$$

$$= \prod_{i=1}^{r} P[\frac{\bar{X}_i - \mu_i}{\sigma/\sqrt{N}} < \frac{d-(\mu_i - \mu_0)}{\sigma/\sqrt{N}}] \prod_{j=k-s+1}^{k} P[\frac{\bar{X}_j - \mu_j}{\sigma/\sqrt{N}} > \frac{d-(\mu_j - \mu_0)}{\sigma/\sqrt{N}}]$$

$$= \prod_{i=1}^{r} \Phi(\frac{\sqrt{N}[d-(\mu_i - \mu_0)]}{\sigma}) \prod_{j=k-s+1}^{k} \Phi(\frac{\sqrt{N}[(\mu_j - \mu_0)-d]}{\sigma}).$$

Since $\mu_i - \mu_0 \leq \delta_1^*$, $\mu_j - \mu_0 \geq \delta_2^*$ and Φ is strictly increasing, it follows that subject to (2.9), the infimum of the PCD is achieved when

$$\text{(2.10)} \qquad \begin{aligned} &\mu_1 = \mu_2 =\ldots= \mu_r = \mu_0 + \delta_1^*, \\ &\mu_{k-s+1} = \mu_{k-s+2} =\ldots= \mu_k = \mu_0 + \delta_2^*; \end{aligned}$$

and under this particular configuration,

$$\text{(2.11)} \qquad P[CD|\vec{\mu}, \sigma^2; R_{1A}] = \Phi^{r+s}(\sqrt{N}/\lambda)$$

where λ is defined in (1.10). Since the r.h.s. of (2.11) depends on r and s only through their sum $(r+s)$ and is decreasing in $(r+s)$, it follows that the r.h.s. of (2.11) is minimized by setting $r+s=k$. This proves (1).

The proof of (2) follows immediately from the fact that for every $\vec{\mu}^0$ specified in (2), we have

$$P[CD|\vec{\mu}^0, \sigma^2; R_{1A}] = \Phi^k(\sqrt{N}/\lambda). \tag{2.12}$$

Theorem 2.1. If N in (1) of Definition 2.1 is the smallest integer satisfying

$$N \geq \lambda^2 z^2 \tag{2.13}$$

where z is such that $\Phi^k_{(z)} = P^*$, then the probability requirement (1.3) is satisfied.

Proof: For every mean vector $\vec{\mu}$,

$$P[CD|\vec{\mu}, \sigma^2; R_{1A}] \geq P[CD|\vec{\mu}^0, \sigma^2; R_{1A}] = \Phi^k(\sqrt{N}/\lambda) \geq \Phi^k(z) = P^*.$$

The most important advantage we enjoy when μ_0 is known is that the statistics $\bar{X}_i - \mu_0$ $(i=1, 2,\ldots, k)$ are independent; so the sample size required to satisfy (1.3) is a function of the percentage point of a univariate normal distribution. When μ_0 is unknown, the statistics $\bar{X}_i - \bar{X}_0$ $(i=1, 2,\ldots, k)$ used in R_{1B} are not independent. It is shown below that the infimum of the PCD under the rule R_{1B} can be expressed in the form of a multivariate normal distribution with mean vector 0 and a certain covariance matrix Σ given in (2.20).

It is easily seen that if a mean vector $\vec{\mu} = \vec{\mu}^0$ is an LF configuration under R_{1B}, the set of populations Ω_I defined in (1.2) must be empty. For every fixed integer $r(0 \leq r \leq k)$, let $\vec{\mu}(r)$ be any mean vector such that there are r populations in Ω_W and $(k-r)$ populations in Ω_B; we can assume without loss of generality that the populations specified by $\vec{\mu}(r)$ are such that

$$\begin{aligned} &\Omega_W = \{\pi_1, \pi_2, \ldots, \pi_r\} \\ &\Omega_B = \{\pi_{r+1}, \pi_{r+2}, \ldots, \pi_k\}. \end{aligned} \tag{2.14}$$

Let $\vec{\mu}^0(r)$ be given by

$$(2.15)\qquad \begin{aligned} &\mu_1^0(r) = \mu_2^0(r) = \dots = \mu_r^0(r) = \mu_0 + \delta_1^*, \\ &\mu_{r+1}^0(r) = \mu_{r+2}^0(r) = \dots = \mu_k^0(r) = \mu_0 + \delta_2^*. \end{aligned}$$

Then it follows that for every $\vec{\mu}(r)$, we have

$$(2.16)\qquad P[CD|\vec{\mu}(r),\ \sigma^2;\ R_{1B}] \geq P[CD|\vec{\mu}^0(r),\ \sigma^2;\ R_{1B}]$$

and

$$P[CD|\vec{\mu}^0(r),\ \sigma^2;\ R_{1B}] = P[\bar{X}_i - \bar{X}_0 \leq d,\ \bar{X}_j - \bar{X}_0 > d\ ({}^{i=1,2,\dots,r}_{j=r+1,r+2,\dots,k})|\vec{\mu}^0(r),\ \sigma^2]$$

$$= P\left[\frac{\bar{X}_i - \mu_i}{\sigma\sqrt{\frac{N}{2}}} - \frac{\bar{X}_0 - \mu_0}{\sigma\sqrt{\frac{N}{2}}} \leq \frac{d-\delta_1^*}{\sigma\sqrt{\frac{N}{2}}},\ \frac{\bar{X}_0 - \mu_0}{\sigma\sqrt{\frac{N}{2}}} - \frac{\bar{X}_j - \mu_j}{\sigma\sqrt{\frac{N}{2}}} \leq \frac{\delta_2^* - d}{\sigma\sqrt{\frac{N}{2}}}\ ({}^{i=1,2,\dots,r}_{j=r+1,r+2,\dots,k})\right].$$

Hence if we define the $(k \times k)$ covariance matrix $\Sigma_r = (\sigma_{ii'})$ by

$$\sigma_{ii'} = \begin{cases} 1 & \text{for } i = i' \\ 1/2 & \text{for } i,\ i' \in \{1,2,\dots,r\} \text{ or } i,\ i' \in \{r+1,\ r+2,\dots,\ k\} \\ -1/2 & \text{for } i \in \{1,2,\dots,r\} \text{ and } i' \in \{r+1,\ r+2,\ \dots,\ k\}; \end{cases}$$

then

$$(2.18)\qquad P[CD|\vec{\mu}^0(r),\ \sigma^2;\ R_{1B}] = \int_{-\infty}^{\sqrt{\frac{N}{2}}/\lambda}\int_{-\infty}^{\sqrt{\frac{N}{2}}/\lambda}\dots\int_{-\infty}^{\sqrt{\frac{N}{2}}/\lambda} \frac{1}{(2\pi)^{k/2}|\Sigma_r|^{1/2}}\, e^{-\frac{1}{2}\underset{\sim}{y}'\Sigma_r^{-1}\underset{\sim}{y}} \prod_{i=1}^{k} dy_i$$

Equation (2.18) gives the infimum of the PCD under the rule R_{1B} for the set of all configurations such that there are r populations in Ω_W and $(k-r)$ populations in Ω_B. To find the LF configuration under R_{1B} it suffices to find the integer where the r.h.s. of (2.18) achieves a minimum over all $r (0 \leq r \leq k)$. We give the LF configuration in the following

<u>Lemma 2.2</u>. For every λ and every N, the LF configuration under R_{1B} is given by $\vec{\mu} = \vec{\mu}^0$ such that

$$(2.19)\qquad \begin{aligned} &\mu_1^0 = \mu_2^0 = \dots = \mu_m^0 = \mu_0 + \delta_1^*, \\ &\mu_{m+1}^0 = \mu_{m+2}^0 = \dots = \mu_k^0 = \mu_0 + \delta_2^* \end{aligned}$$

where m is defined in (1.7).

Proof: By Theorem A.1 in Appendix A, the r.h.s. of (2.18) is minimized at $r = m$. Combining this result with (2.16) gives the desired result, i.e.,

$$\inf_{\vec{\mu}} P[CD|\vec{\mu}, \sigma^2; R_{1B}] = \min_{r} \inf_{\vec{\mu}(r)} P[CD|\vec{\mu}(r), \sigma^2; R_{1B}]$$

$$= \min_{r} P[CD|\vec{\mu}^0(r), \sigma^2; R_{1B}] = P[CD|\vec{\mu}^0, \sigma^2; R_{1B}].$$

Remark 2.3. It is important to observe that for both R_{1A} and R_{1B}, the LF configuration does not depend on σ^2 and the sample size N. This property is used in Chapters III and IV where σ^2 is unknown.

Now let Σ denote the $(k \times k)$ positive definite covariance matrix $\Sigma_{1,2}$ defined in (2.17), i.e., Σ has the following structure:

$$(2.20) \qquad \Sigma = \begin{array}{c} \begin{array}{cc} 1\ 2 \ldots m & m+1 \ldots k \end{array} \\ \left(\begin{array}{c|c} \begin{matrix} 1 & & \frac{1}{2} \\ & \ddots & \\ \frac{1}{2} & & 1 \end{matrix} & -\frac{1}{2} \\ \hline -\frac{1}{2} & \begin{matrix} 1 & & \frac{1}{2} \\ & \ddots & \\ \frac{1}{2} & & 1 \end{matrix} \end{array}\right) \end{array} \begin{array}{c} 1 \\ 2 \\ \vdots \\ m \\ m+1 \\ \vdots \\ k \end{array}$$

Let $b = b(P, k)$ be the solution of the equation

$$(2.21) \qquad P = \int_{-\infty}^{b} \int_{-\infty}^{b} \cdots \int_{-\infty}^{b} \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} e^{-\frac{1}{2} \underset{\sim}{y}'\Sigma^{-1}\underset{\sim}{y}} \prod_{i=1}^{k} dy_i.$$

Then the sample size N required under R_{1B} can be determined by the following

Theorem 2.2. Let λ be defined in (1.10) and b be the solution of (2.21) with $P = P^*$. If N in (2) of Definition 2.1 is the smallest integer satisfying

$$(2.22) \qquad N \geq 2\lambda^2 b^2$$

then the probability requirement (1.3) is satisfied.

Proof: Using Lemma 2.2, (2.18) and (2.22) we have

$$\inf_{\vec{\mu}} P[CD|\vec{\mu}, \sigma^2; R_{1B}] = P[CD|\vec{\mu}^0, \sigma^2; R_{1B}]$$

$$= \int_{-\infty}^{\sqrt{\frac{N}{2}}/\lambda} \int_{-\infty}^{\sqrt{\frac{N}{2}}/\lambda} \cdots \int_{-\infty}^{\sqrt{\frac{N}{2}}/\lambda} \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} e^{-\frac{1}{2} \underset{\sim}{y}'\Sigma^{-1}\underset{\sim}{y}} \prod_{i=1}^{k} dy_i \geq P^*.$$

The solution $b = b(P, k)$ of (2.21) is the equi-coordinate percentage point of a k-dimensional multivariate normal distribution with mean vector 0 and covariance matrix Σ given in (2.20). The values of b as a function of P and k have been tabulated in Table 1 which can be found at the end of this work for $P = 0.50, 0.75, 0.90, 0.95, 0.975, 0.99$ and $k = 1(1)10(2)20$. It should be noted that for $k = 1$ the table reduces to the univariate standard normal table. The numerical solution was obtained by first obtaining another expression for the result. Let $Z_0, Z_1, \ldots, Z_k$ be independent standard normal chance variables. For $\varphi(\cdot)$, $\Phi(\cdot)$ and m given in (1.5), (1.6) and (1.7) respectively, we define

$$(2.23) \qquad Y_i = \begin{cases} (Z_i - Z_0)/\sqrt{2} & \text{for } i \leq m, \\ (Z_0 - Z_i)/\sqrt{2} & \text{for } i > m. \end{cases}$$

Then $(Y_1, Y_2, \ldots, Y_k)$ follows a multivariate normal distribution with mean vector 0 and covariance matrix Σ. It follows that the b-value satisfying (2.21) also satisfies

$$P = P[Y_i \leq b \ (i=1, 2, \ldots, k)]$$

$$= P[Z_i \leq Z_0 + \sqrt{2}b, \ Z_j > Z_0 - \sqrt{2}b \ (\begin{smallmatrix} i=1, 2, \ldots, m \\ j=m+1, m+2, \ldots, k \end{smallmatrix})]$$

or

$$(2.24) \qquad P = \int_{-\infty}^{\infty} \Phi^m(z + \sqrt{2}b)\Phi^{k-m}(-z + \sqrt{2}b)\varphi(z)dz,$$

and vice versa.

The computation of Table 1 was based on equation (2.24) and was carried out on a CDC 6600 computer at the University of Minnesota. The r.h.s. integral in (2.24) was approximated by a Gaussian quadrature summation formula given in [31]. To be conservative, the entries in the table have all been rounded to the next higher value (in the 7th decimal) and should be in error by at most one unit in the last digit given.

2.2 An Upper Bound on the Sample Size Required.

In this Section we g--- an upper bound on the sample size N under R_{1B} by using the following

Lemma 2.3. For any given $P \epsilon [0, 1]$ and any two events A and B,

$$P(A) + P(B) = 1 + P \Longrightarrow P[A \cap B] \geq P, \tag{2.25}$$

and the equality holds iff $P[A \cup B] = 1$.

Proof: It is an immediate consequence of the inequality

$$P[A \cap B] = P[(A^c \cup B^c)^c] \geq 1 - [P(A^c) + P(B^c)] = P(A) + P(B) - 1$$

which follows from Boole's inequality (see, e.g., [13: p.23]); and the equality follows iff $P(A^c \cap B^c) = 0$ or $P(A \cup B) = 1$.

Now for any real number c and positive integer q, we define

$$H_q(c) = \int_{-\infty}^{c} \int_{-\infty}^{c} \cdots \int_{-\infty}^{c} \frac{1}{(2\pi)^{q/2} |\Sigma'|^{1/2}} e^{-\frac{1}{2} \underset{\sim}{y}'(\Sigma')^{-1} \underset{\sim}{y}} \prod_{i=1}^{q} dy_i \tag{2.26}$$

where the $q \times q$ convariance matrix $\Sigma' = (\sigma_{ij})$ is such that

$$\sigma_{ij} = \begin{cases} 1 & \text{if } i = j \\ \frac{1}{2} & \text{if } i \neq j. \end{cases} \tag{2.27}$$

Let b be the solution of equation (2.21) and b' be the solution of the equation

$$H_m(b') + H_{k-m}(b') = 1 + P. \tag{2.28}$$

Theorem 2.3. For every P and every k, we have

$$b' > b. \tag{2.29}$$

Proof: Let $(Y_1, Y_2, \ldots, Y_k)$ follow a multivariate normal distribution with mean vector 0 and covariance matrix Σ, and let

$$A = [Y_i \leq b' \quad (i = 1,2,\ldots, m)], \tag{2.30}$$

$$B = [Y_i \leq b' \quad (i = m+1, m+2, \ldots, k)]; \tag{2.31}$$

then $H_m(b') = P(A)$ and $H_{k-m}(b') = P(B)$. By Lemma 2.3,

$$H_m(b') + H_{k-m}(b') = 1 + P \Longrightarrow P[A \cap B] = P[Y_i \leq b'(i=1,2,\ldots,k)] > P.$$

It follows that $b' > b$ and this completes the proof.

Corollary. Let N be defined in (2.22). If N' is the smallest integer satisfying

$$N' \geq 2\lambda^2 b'^2 \tag{2.32}$$

where b' is the solution of (2.28) with $P = P^*$, then $N^* \geq N$.

When k is even, equation (2.28) reduces to

$$H_{k/2}(b') = \frac{1}{2}(1 + P). \tag{2.33}$$

The solution b' of (2.33) is the percentage point of an equicorrelated multivariate normal distribution. The numerical solutions have been tabulated by both Gupta [15] and Milton [22] at several probability levels. Let $\gamma = \gamma(P^*, k)$ denote the quantity $(b'/b)^2$ with $P = P^*$, then we have approximately

$$N'/N \approx \gamma. \tag{2.34}$$

These γ values have been computed for even k based on the b' values given by Milton and the b values given in Table 1 of this work, an excerpt is given below:

Values of γ for Selected Values of P^* and k

P^* \ k	2	4	6	8	12	16
0.50	1.102597	1.040598	1.026127	1.019653	1.013570	1.010628
0.90	1.000353	1.000126	1.000071	1.000047	1.000027	1.000018
0.95	1.000040	1.000013	1.000007	1.000005	1.000002	1.000002

The computation shows the bound given in (2.32) is quite good for most purposes since most of the γ values are very close to 1. Of course, the value of $b'/b = \sqrt{\gamma}$ is even closer to one. It also appears that $\gamma(P^*, k)$ is monotone decreasing both in P^* and in k; however, the author has not tried to prove this result.

2.3 Some Optimal Properties of the Decision Rule.

In this section we prove some of the optimal properties of the proposed decision rule specified in Definition 2.1. We will restrict our discussion to R_{1B} only, and an analogous discussion of all of this section can be given for R_{1A}.

Lehmann [18] in 1957 considered the theoretical aspects of a class of multiple decision procedures and proved several fundemental theorems. It is seen below that our problem can be fitted within this framework. For this purpose we first give the following

Definition 2.3. For $i = 1,2,\ldots, k$, let (ψ_i, D_i, L_i) be k component statistical decision problems where ψ_i is the parameter space, D_i is the decision space and $L: \psi_i \times D_i \to (-\infty, \infty)$ is the loss function for the i-th component decision problem. The decision problem (ψ, D, L) is said to be the corresponding product decision problem if

(1) $\psi = \underset{i=1}{\overset{k}{X}} \psi_i = \{\vec{\mu} = (\mu_1, \mu_2,\ldots, \mu_k): \mu_i \epsilon \psi_i,\ i = 1,2,\ldots, k\}$ is the product parameter space,

(2) $D = \underset{i=1}{\overset{k}{X}} D_i = \{\vec{d} = (d_1, d_2,\ldots, d_k): d_i \epsilon D_i,\ i = 1,2,\ldots, k\}$ is the product decision space, and

(3) $L = L(\vec{\mu}, \vec{d})$ is the loss function defined on $\psi \times D$.

Remark 2.4. The problem of incompatibility of two component decision rules, which was discussed by Lehmann, does not arise in our formulation; hence we omit it.

Now for $i = 1,2,\ldots, k$, let (ψ_i, D_i, L_i) be the component decision problem dealing with the population mean of π_i. Throughout this section, we will use the simple loss function; i.e., for any component decision rule $d_i \epsilon D_i$, the loss function $L_i = L(\mu_i, d_i)$ is such that

decisions	states of nature		
	$\pi_i \epsilon \Omega_W$	$\pi_i \epsilon \Omega_I$	$\pi_i \epsilon \Omega_B$
$\pi_i \epsilon S_W$	0	0	1
$\pi_i \epsilon S_B$	1	0	0

or equivalently, with the term "misclassification" defined in Definition 1.2, (2.35) $\quad L(\mu_i, d_i) = \begin{cases} 1 & \text{if } \pi_i \text{ is misclassified,} \\ 0 & \text{otherwise;} \end{cases}$

and the corresponding risk function is

(2.36) $\quad \gamma(\mu_i, d_i) = P[\pi_i \text{ is miscallsified } |\mu_i, \sigma^2; d_i].$

Let (ψ, D, L) be the product decision problem of partitioning the set of k populations into two disjoint subsets S_W and S_B as stated in Chapter I. For any decision rule $\vec{d} = (d_1, d_2, \ldots, d_k)$ and true parameter value $\vec{\mu}$, if the following loss function

(2.37) $\quad L(\vec{\mu}, \vec{d}) = M$

is used where M is the number of populations misclassified, then

(2.38) $\quad \gamma(\vec{\mu}, \vec{d}) = E\, L(\vec{\mu}, \vec{d}) = E\, M$

is the expected misclassification size when the mean vector is $\vec{\mu}$ and when the decision rule $\vec{d}$ is used.

It should be observed that the component loss functions $L_i = L(\mu_i, d_i)$ $(i = 1,2,\ldots, k)$ and $L(\vec{\mu}, \vec{d})$ are related by

(2.39) $\quad L(\vec{\mu}, \vec{d}) = \sum_{i=1}^{k} L(\mu_i, d_i);$

it follows from this that

(2.40) $\quad \gamma(\vec{\mu}, \vec{d}) = \sum_{i=1}^{k} \gamma(\mu_i, d_i).$

For Φ defined in (1.6) and b defined in (2.21) with $P = P^*$, the following lemma gives an upper bound on the risk under R_{1B}.

Lemma 2.4. Under the rule R_{1B},

(1) $EM = EM(\vec{\mu}, \sigma^2) \leq k[1 - \Phi(b)]$ for every $\vec{\mu}$;

(2) the equality holds when μ_i is either $\mu_0 + \delta_1^*$ or $\mu_0 + \delta_2^*$ for each i. In particular, it holds at $\vec{\mu} = \vec{\mu}^0$ where $\vec{\mu}^0$ is the LF configuration defined in (2.19).

Proof: Let Z be a standard normal chance variable. Then for every $\vec{\mu}$,

$$
\begin{aligned}
EM &= \sum_{\{i:\pi_i \in \Omega_W\}} P[\bar{X}_i - \bar{X}_0 > d \mid \mu_i, \sigma^2] + \sum_{\{j:\pi_j \in \Omega_B\}} P[\bar{X}_j - \bar{X}_0 < d \mid \mu_j, \sigma^2] \\
&\leq \sum_{\{i:\pi_i \in \Omega_W\}} P[\bar{X}_i - \bar{X}_0 > d \mid \mu_i = \mu_0 + \delta_1^*, \sigma^2] + \sum_{\{j:\pi_j \in \Omega_B\}} P[\bar{X}_j - \bar{X}_0 < d \mid \mu_j = \mu_0 + \delta_2^*, \sigma^2] \\
&\leq kP[Z > \sqrt{\tfrac{N}{2}}/\lambda] = k[1 - \Phi(b)]
\end{aligned}
$$

where the last equality follows from (2.22), if we disregard the fact that N has to be an integer. This proves (1).

The proof of (2) follows immediately from the fact that all inequalities in the proof of (1) are equalities when μ_i is either $\mu_0 + \delta_1^*$ or $\mu_0 + \delta_2^*$.

The next lemma was given by Lehmann [18]; although the original statement was for $k = 2$ only, it follows from the proof that the conclusion holds for any positive integer k.

Lemma 2.5. (Lehmann). For $i = 1,2,\ldots, k$, let ρ_i be the a priori distribution on ψ_i and $\delta_i = \delta_i(\rho_i)$ be a Bayes decision rule for the component decision problem (ψ_i, D_i, L_i) with loss function $L(\mu_i, d_i)$; and let ρ be the product a priori probability measure defined on $\psi = \overset{k}{\underset{i=1}{X}} \psi_i$. If the loss function $L(\vec{\mu}, \vec{d})$ for the product decision problem (ψ, D, L) is such that

$$
L(\vec{\mu}, \vec{d}) = \sum_{i=1}^{k} c_i L(\mu_i, d_i), \quad c_i > 0 \quad (i = 1,2,\ldots, k), \tag{2.41}
$$

then the product decision rule $\vec{\delta} = (\delta_1, \delta_2, \ldots, \delta_k)$ is a Bayes decision rule w.r.t. ρ for the product decision problem.

Remark 2.5. It also follows from the proof that if δ_i is the unique Bayes decision rule w.r.t. ρ_i for (ψ_i, D_i, L_i) $(i = 1, 2,\ldots, k)$, then $\vec{\delta}$ is the unique Bayes decision rule w.r.t. ρ for (ψ, D, L).

By (2.39), Lemma 2.5 implies that in order to show the decision rule specified by R_{1B} is Bayesian w.r.t. ρ, it suffices to show that for every $i = 1,2,\ldots, k$, the component decision rule

$$
\begin{cases} \pi_i \in S_W & \text{if } \bar{X}_i - \bar{X}_0 < d, \\ \pi_i \in S_B & \text{if } \bar{X}_i - \bar{X}_0 > d \end{cases} \tag{2.42}
$$

is Bayesian w.r.t. ρ_i for the decision problem dealing with μ_i when the loss function (2.35) is used. For this purpose we prove the following

Lemma 2.6. Let (ψ_i, D_i, L_i) be a statistical decision problem such that ψ_i contains two density functions, i.e., $\psi_i = \{f_1, f_2\}$. When observation x is taken, the decision rule is to define a set W such that we assert the state of nature is f_2 iff $x \epsilon W$. Let

$$\alpha = P[x \epsilon W | f_1],$$

$$\beta = P[x \not\epsilon W | f_2].$$

Then $\alpha + \beta$ is minimized by taking $W = W_0$ defined by

$$W_0 = \{x: \frac{f_2(x)}{f_1(x)} > 1\}. \tag{2.43}$$

Proof: $\alpha + \beta = \int_W f_1(x)dx + \int_{W^c} f_2(x)dx = 1 - \int_W [f_2(x) - f_1(x)]dx$, the conclusion follows from the fact that the integral on the r.h.s. is maximized when $W = W_0$.

Corollary. If f_1 and f_2 are density functions of normal variables with means δ_1^* and δ_2^* respectively and a common known variance σ_0^2, then the set W_0 defined in (2.43) reduces to:

$$W_0 = \{x: x > d\} \tag{2.44}$$

where $d = \frac{1}{2}(\delta_1^* + \delta_2^*)$. It should be noted that W_0 does not depend on σ_0^2.

Now for $i = 1, 2, \ldots, k$, let the a priori distribution of ψ_i be

$$\rho_i(\mu_i) = \begin{cases} \frac{1}{2} & \text{for } \mu_i = \mu_0 + \delta_1^*, \\ \frac{1}{2} & \text{for } \mu_i = \mu_0 + \delta_2^* \end{cases} \tag{2.45}$$

and $\rho(\vec{\mu})$ be the product a priori probability measure defined on ψ. Then we are able to prove the following

Theorem 2.4. The decision rule specified by R_{1B} is the unique Bayes solution with respect to the a priori distribution ρ.

Proof: Let d_i be any component decision rule for (ψ_i, D_i, L_i) based on the sufficient statistics $\overline{X}_i - \overline{X}_0$ and the region W, where d_i asserts that

$\pi_i \epsilon S_B$ if and only if $(\overline{X}_i - \overline{X}_0)\epsilon W$. Then its corresponding risk function is

$$(2.46) \qquad \gamma_i(d_i) = \frac{1}{2}\{P[(\overline{X}_i-\overline{X}_0)\epsilon W|\mu_i=\mu_0+\delta_1^*] + P[(\overline{X}_i-\overline{X}_0)\notin W|\mu_i=\mu_0+\delta_2^*]\}.$$

By the Corollary of Lemma 2.6, the only component decision rule δ_i which minimizes (2.46) is the one in which $W \doteq W_0$, where W_0 is given in (2.44). This shows that the component decision rule δ_i defined in (2.42) is the unique Bayes solution for the component decision problem (ψ_i, D_i, L_i) with respect to ρ_i, and hence by Lemma 2.5 the rule $\vec{\delta}$ specified by R_{1B} is the unique Bayes solution w.r.t. ρ for the product decision problem.

The following lemma is also due to Lehmann [20: p 4-19].

Lemma 2.7. (Lehmann). Let X have distribution $P_\theta(x)$, $\theta\epsilon\psi$. Suppose there is a distribution ρ over ψ and a set $\omega \subset \psi$ such that $\rho(\omega) = 1$ and

$$(2.47) \qquad \gamma_{\vec{\delta}}(\theta) = \sup_{\theta'\epsilon\psi} \gamma_{\vec{\delta}}(\theta') \qquad \text{for all } \theta\epsilon\omega.$$

Then we can conclude that

(1) $\vec{\delta} = \vec{\delta}_\rho$ is minimax, and

(2) ρ is least favorable.

Theorem 2.5. The decision rule specified by R_{1B} is minimax and admissible.

Proof: For ρ_i defined in (2.45), let ρ be the corresponding product a priori probability measure defined on ψ; then either $\mu_i = \mu_0 + \delta_1^*$ or $\mu_i = \mu_0 + \delta_2^*$ a.s. By Lemma 2.4, the condition (2.47) is satisfied. Hence the decision rule $\vec{\delta}$ specified by R_{1B} is minimax.

The admissibility follows from Theorem 2.4 and the fact that an unique Bayes decision rule is admissible.

CHAPTER III A Two-Stage Procedure

3.1 Introduction.

In the previous chapter we have treated the problem in which the common population variance σ^2 is known. An explicit rule has been given and, it should be noted that the sample size required to satisfy the probability requirement is a linear function of σ^2 (see (2.13) and (2.22)). A natural outgrowth of this problem is to ask what sampling procedure and decision rule should be used to solve the problem if σ^2 is unknown. This is more realistic because the true value of σ^2 is not available in most applied problems.

It is easily seen that when σ^2 is unknown, there does not exist a decision rule which can solve the present problem under any single-stage sampling procedure; a proof analogous to Dantzig's non-existence theorem [9] can be obtained easily. In 1945 Stein [30] suggested a two-stage procedure to provide a solution for the problem considered by Dantzig; later Bechhofer, Dunnett and Sobel [4] employed the idea of Stein's two-stage procedure to solve the problem of selecting the population associated with the largest mean in k normal populations with a common unknown variance. All of this suggests that the present problem can also be solved under a two-stage procedure when σ^2 is unknown. The proposed rule is given in Section 3.2. When μ_0 is unknown, the determination of the sample size required depends on the percentage points of a multivariate generalization of Student's t distribution. The property of this distribution is investigated in Section 3.3 and the percentage points of this distribution as a function of P, k and degrees of freedom ν have been computed and tabulated in Table 2, which can be found at the end of this work. The expected sample size and expected misclassification size are investigated in Sections 3.4 and 3.5, respectively. The relative efficiency of the proposed rule is discussed in Section 3.4.

3.2 The Proposed Rule.

Let w_ν be the solution of the equation

$$(3.1) \qquad F_\nu^k(w_\nu) = P^*$$

where $F_\nu(\cdot)$ is the c.d.f. of Student's t-distribution with ν degrees of freedom. For μ_0 known, the following rule is proposed:

Definition 3.1. Rule R_{2A}:

(1) Let $n_0 \geq 2$ be a preassigned positive integer. We observe the sequence defined in (2.1) for $j = 1, 2, \ldots, n_0$. Compute

$$S_\nu^2 = \frac{1}{\nu} \sum_{i=1}^{k} \sum_{j=1}^{n_0} [X_{ij} - \frac{1}{n_0} (\sum_{j=1}^{n_0} X_{ij})]^2$$

with $\nu = k(n_0 - 1)$.

(2) Observe the sequence defined in (2.1) for $j = n_0 + 1, n_0 + 2, \ldots, N$ where the total sample size N is the smallest integer satisfying

$$(3.2) \qquad N \geq \max \{n_0, S_\nu^2 w_\nu^2/a^2\}$$

with a defined in (1.9).

(3) Compute the k overall sample means

$$(3.3) \qquad \bar{X}_i = \frac{1}{N} \sum_{j=1}^{N} X_{ij} \quad \text{for } i = 1, 2, \ldots, k$$

and apply the decision rule defined in (2.5).

Lemma 3.1. For every $\vec{\mu}$ and every σ^2,

$$(3.4) \qquad P[CD|\vec{\mu}, \sigma^2; R_{2A}] \geq P^*.$$

Proof: Since the LF configuration $\vec{\mu}^0$ given in Lemma 2.1 does not depend on σ^2 and N (see Remark 2.3), it follows that we can restrict our attention to $\vec{\mu} = \vec{\mu}^0$ in (3.4).

Let t be Student's t variable with ν degrees of freedom. Then

$$P[CD|\vec{\mu}^0, \sigma^2; R_{2A}] = \prod_{i=1}^{k} P[\frac{\sqrt{N}(\bar{X}_i - \mu_i^0)/\sigma}{S_\nu/\sigma} \leq \frac{\sqrt{N}\, a}{S_\nu}] = P^k[t \leq \frac{\sqrt{N}\, a}{S_\nu}] \geq P^*$$

where the last inequality (which is independent of σ^2) follows from the fact that for any observed S_ν in the first stage, we have $\sqrt{N}\, a/S_\nu \geq w_\nu$ from (3.2).

When μ_0 is unknown, we propose the following rule:

Definition 3.2. Rule R_{3B}:

(1) Let $n_0 \geq 2$ be a preassigned positive integer. We observe the sequence defined in (2.3) for $j = 1, 2, \ldots, n_0$. Compute

$$S_\nu^2 = \frac{1}{\nu} \sum_{i=0}^{k} \sum_{j=1}^{n_0} [X_{ij} - \frac{1}{n_0} (\sum_{j=1}^{n_0} X_{ij})]^2$$

with $\nu = (k+1)(n_0-1)$.

(2) Observe the sequence defined in (2.3) for $j = n_0+1, n_0+2, \ldots, N$ where N is to be determined below.

(3) Compute the $(k+1)$ overall sample means

$$(3.5) \qquad \bar{X}_i = \frac{1}{N} \sum_{j=1}^{N} X_{ij} \qquad \text{for } i = 0, 1, \ldots, k$$

and apply the decision rule defined in (2.6).

To determine the sample size N in the above rule, we first observe that the LF configuration $\vec{\mu}^0$ (given in (2.19)) does not depend on σ^2 and N. Hence to satisfy the probability requirement (1.4), we can again restrict our attention to $\vec{\mu}^0$. Let $(Y_1, Y_2, \ldots, Y_k)$ follow a multivariate normal distribution with mean vector 0 and covariance matrix Σ defined in (2.20), and let $U_\nu = \frac{S_\nu}{\sigma}$. Then

$$P[CD|\vec{\mu}^0, \sigma^2; R_{2B}] = P[\bar{X}_i - \bar{X}_0 < d, \bar{X}_j - \bar{X}_0 > d \, (\substack{i=1, 2, \ldots, m \\ j=m+1, m+2, \ldots, k})|\vec{\mu}^0]$$

$$= P[\frac{Y_i}{U_\nu} \leq \sqrt{\frac{N}{2}} a/S_\nu \quad (i = 1, 2, \ldots, k)]$$

$$= P[t_i \leq \sqrt{\frac{N}{2}} a/S_\nu \quad (i = 1, 2, \ldots, k)|\nu]$$

where $t_1, t_2, \ldots, t_k$ are Student's t variables with ν degrees of freedom each, and they are correlated with correlation matrix Σ. The joint distribution of $(t_1, t_2, \ldots, t_k)$ follows a multivariate t distribution which will be discussed in Section 3.3.

Let h_ν be such that

$$(3.6) \qquad P[t_i \leq h_\nu \, (i = 1, 2, \ldots, k)|\nu] = P^*$$

(for a mathematical expression of this equation, see (3.12)). Then the sample size N can be determined in the following

Theorem 3.1. If N in Definition 3.2 is the smallest integer satisfying

(3.7) $\quad N \geq \max \{n_0, 2S_\nu^2 h_\nu^2/a^2\}$,

with a defined in (1.9), then the probability requirement (1.4) is satisfied.

Proof: $\quad P[CD|\vec{\mu}, \sigma^2; R_{2B}] \geq P[CD|\vec{\mu}^0, \sigma^2; R_{2B}] = P[t_i \leq \sqrt{\frac{N}{2}}\, a/S_\nu \ (i=1,2,\ldots,k)|\nu] \geq P^*$

where the last inequality follows from the fact that for any observed S_ν in the first stage, we have $\sqrt{\frac{N}{2}}\, a/S_\nu \geq h_\nu$ from (3.7).

3.3 Properties of the Multivariate t Distribution.

Let $(Y_1, Y_2,\ldots, Y_k)$ follow a multivariate normal distribution with mean vector 0 and covariance matrix Σ, and let U_ν be such that νU_ν^2 follows a chi-square distribution with ν degrees of freedom and U_ν is independent of $(Y_1, Y_2,\ldots, Y_k)$. We assume Σ has the form defined in (2.20). However, it is easily seen that all of the results in this section also apply to any covariance matrix with $\sigma_{ii} = 1$ $(i = 1, 2,\ldots, k)$.

If we define

$$t_i = \frac{Y_i}{U_\nu} \quad \text{for } i = 1, 2,\ldots, k \tag{3.8}$$

then t_i is a Student's t variable with ν degrees of freedom for every i and the random vector $(t_1, t_2,\ldots, t_k)$ is said to follow a multivariate t distribution with correlation matrix Σ.

The purpose of this section is to derive the joint distribution of $(t_1,t_2,\ldots,t_k)$ and to examine a property of the equi-coordinate percentage points of this distribution. Some results in this area with a different covariance matrix were given by other authors; in particular see [11] and [16].

Since for given $U_\nu = u$ $(t_1, t_2,\ldots, t_k)$ has a multivariate normal distribution with mean vector 0 and covariance matrix $\frac{1}{u^2}\Sigma$, it follows that the joint density function of $(t_1, t_2,\ldots, t_k)$ is given by

$$f_{k,\nu,\Sigma}(t_1, t_2,\ldots, t_k) = \int_0^\infty \frac{1}{(2\pi)^{k/2}|\frac{1}{u^2}\Sigma|^{1/2}}\, e^{-\frac{1}{2}\underset{\sim}{t}'(\frac{1}{u^2}\Sigma)^{-1}\underset{\sim}{t}}\, \gamma_\nu(u)du \tag{3.9}$$

where

$$\gamma_\nu(u) = \begin{cases} \dfrac{2(\frac{\nu}{2})^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})}\, u^{\nu-1} e^{-\frac{\nu}{2}u^2} & \text{for } u > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.10}$$

is the density function of U_ν (see, e.g., [8: p. 237]). Let $s = \frac{1}{2}(\underset{\sim}{t}'\Sigma^{-1}\underset{\sim}{t} + \nu)u^2$ in (3.9). Then

$$f_{k,\nu,\Sigma}(t_1,t_2,\ldots,t_k) = \frac{2(\frac{\nu}{2})^{\frac{\nu}{2}}}{(2\pi)^{k/2}|\Sigma|^{1/2}\Gamma(\frac{\nu}{2})} \int_0^\infty 2^{\frac{k+\nu}{2}-1} (\underset{\sim}{t}'\Sigma^{-1}\underset{\sim}{t}+\nu)^{-\frac{k+\nu}{2}} e^{-s} s^{\frac{k+\nu}{2}-1} ds$$

or

$$(3.11)\quad f_{k,\nu,\Sigma}(t_1,t_2,\ldots,t_k) = \frac{\Gamma(\frac{k+\nu}{2})}{(\nu\pi)^{k/2}|\Sigma|^{1/2}\Gamma(\frac{\nu}{2})} [1 + \frac{1}{\nu}\underset{\sim}{t}'\Sigma^{-1}\underset{\sim}{t}]^{-\frac{k+\nu}{2}}$$

for $t_i \varepsilon (-\infty, \infty)$, $i = 1, 2,\ldots, k$.

The density function given in (3.11) in general depends on k, ν and Σ where Σ is defined in (2.20). It should be observed that when $k = 1$, it reduces to the density of a Student's t variable with ν degrees of freedom; and when $\nu \to \infty$ it converges to the joint density of a k-dimensional multivariate normal chance variables with mean 0 and covariance matrix Σ.

For every given ν, let $(t_1, t_2,\ldots, t_k)$ have the joint density function defined in (3.11), and let $h_\nu = h_\nu(P, k)$ be the solution of the equation

$$(3.12)\quad \begin{aligned} P &= P[t_i \le h_\nu\ (i = 1, 2,\ldots, k)|\nu] \\ &= \frac{\Gamma(\frac{k+\nu}{2})}{(\nu\pi)^{k/2}|\Sigma|^{1/2}\Gamma(\frac{\nu}{2})} \int_{-\infty}^{h_\nu}\int_{-\infty}^{h_\nu}\cdots\int_{-\infty}^{h_\nu} [1 + \frac{1}{\nu}\underset{\sim}{t}'\Sigma^{-1}\underset{\sim}{t}]^{-\frac{k+\nu}{2}} \prod_{i=1}^{k} dt_i. \end{aligned}$$

Then h_ν is defined to be the equi-coordinate percentage point of $(t_1, t_2,\ldots, t_k)$. In the following theorem we investigate the behavior of h_ν when ν is large.

Theorem 3.2. For every $P\varepsilon(0, 1)$, let b satisfy (2.21) and h_ν satisfy (3.12). Then

$$(3.13)\quad h_\nu \to b \text{ as } \nu \to \infty.$$

Proof: Let c be any real number and let

$$(3.14)\quad Y = \max_{1\le i\le k} Y_i,$$

$$(3.15)\quad t = \max_{1\le i\le k} t_i;$$

then it is easily seen that

$$(3.16)\quad P[t_i \le c\ (i = 1, 2,\ldots, k)|\nu] = P[t \le c|\nu] = P[\frac{Y}{U_\nu} \le c].$$

We consider the sequence of chance variables $\{\frac{Y}{U_\nu}\}$. Since U_ν^2 is distributed as $\frac{1}{\nu}\sum_{i=1}^{\nu} V_i$ where the V's are i.i.d. chi-square variables with 1 degree of freedom, it follows from the strong law of large numbers that $U_\nu^2 \to 1$ a.s., which implies $U_\nu \to 1$ a.s. (since $U_\nu > 0$ a.s.). By a convergence theorem in Cramer [8: p. 254], we have

$$\frac{Y}{U_\nu} \xrightarrow{d} Y \quad \text{as} \quad \nu \to \infty. \tag{3.17}$$

This shows that $P[t \leq c|\nu] \to P[Y \leq c]$ as $\nu \to \infty$.

The rest of the argument follows from Theorem B.1 given in Appendix B.

The values of h_ν have been computed and tabulated in Table 2 of this work for $P = 0.50, 0.75, 0.90, 0.95, 0.975$ and 0.99; $k = 2(1)6(2)12(4)20$; and $\nu = 5(1)10(2)20(4)60(30)120$. The corresponding b values given in Table 1 is repeated here under $\nu = \infty$. It should be noted that for $k = 1$ the Student's t table can be used.

The following method was used for the computation of Table 2: First we observe that

$$P = P[t_i \leq h_\nu \ (i = 1, 2, \ldots, k)|\nu] = EP[Y_i \leq U_\nu h_\nu \ (i = 1, 2, \ldots, k)]$$

$$= \int_0^\infty [\int_{-\infty}^\infty \Phi^m(z + \sqrt{2}\, uh_\nu)\Phi^{k-m}(-z + \sqrt{2}\, uh_\nu)\varphi(z)dz]\gamma_\nu(u)du$$

where the expection is over U_ν space, the last equality follows from (2.24) and $\gamma_\nu(u)$ is defined in (3.10).Substituting for $y = \frac{\nu}{2}u^2$ in the integration. we have

$$P = \int_0^\infty [\int_{-\infty}^\infty \Phi^m(z + 2h_\nu\sqrt{\frac{y}{\nu}})\Phi^{k-m}(-z + 2h_\nu\sqrt{\frac{y}{\nu}})\varphi(z)dz]\ \frac{1}{\Gamma(\frac{\nu}{2})}\ y^{\frac{\nu}{2}-1} e^{-y}dy. \tag{3.18}$$

Hence if h_ν satisfies (3.12), it must satisfy (3.18) and vice versa. The computation of Table 2 was based on (3.18) and was carried out on a CDC 6600 computer at the University of Minnesota. The double integral on the r.h.s. of (3.18) was approximated by a double summation based on Gaussian quadrature formula given in [31]. To be conservative, the entries in Table 2 have all been rounded to the next higher value (in the 5th decimal) and should be in error by at most one unit in the last digit given.

3.4 The Expected Sample Size and Relative Efficiency.

In this section the expected sample size and its asymptotic behavior of the two-stage procedure will be investigated. We will restrict our attention to R_{2B} only; the argument for R_{2A} is similar and is omitted.

Let N be the random sample size defined in (3.7). It is easily seen that

$$(3.19) \qquad P[N = n] = \begin{cases} 0 & \text{for } n < n_0, \\ P[2S_\nu^2 h_\nu^2/a^2 \le n_0] & \text{for } n = n_0, \\ P[n - 1 < 2S_\nu^2 h_\nu^2/a^2 \le n] & \text{for } n \ge n_0 + 1. \end{cases}$$

Denote by

$$(3.20) \qquad r = \frac{1}{2\lambda^2 h_\nu^2},$$

since $\nu S_\nu^2/\sigma^2$ is a chi-square variable with ν degrees of freedom, (3.19) can be rewritten as

$$(3.21) \qquad P[N = n] = \begin{cases} 0 & \text{for } n < n_0, \\ P[\chi_\nu^2 \le r\nu n_0] & \text{for } n = n_0, \\ P[r\nu(n-1) < \chi_\nu^2 \le r\nu n] & \text{for } n \ge n_0 + 1. \end{cases}$$

Hence

$$(3.22) \qquad EN = n_0 P[N = n_0] + \sum_{n=n_0+1}^{\infty} nP[N = n]$$

$$= n_0 P[\chi_\nu^2 \le r\nu n_0] + \sum_{n=n_0+1}^{\infty} \frac{n}{2^{\nu/2}\Gamma(\frac{\nu}{2})} \int_{r\nu(n-1)}^{r\nu n} y^{\frac{\nu}{2} - 1} e^{-\frac{y}{2}} dy.$$

Consider any fixed summand in the second term on the r.h.s. of (3.22). Since for $r\nu(n-1) \le y \le r\nu n$ n satisfies

$$(3.23) \qquad y/r\nu \le n \le y/r\nu + 1,$$

using the first inequality in (3.23), the second term Q on the r.h.s. of (3.22) can be bounded by

(3.24) $$Q \geq \sum_{n=n_0+1}^{\infty} \frac{1}{2^{\nu/2}\Gamma(\frac{\nu}{2})} \int_{r\nu(n-1)}^{r\nu n} \frac{1}{r\nu}\, y^{\frac{\nu}{2}} e^{-\frac{y}{2}}\, dy = \frac{1}{r} P[\chi^2_{\nu+2} > r\nu n_0];$$

similarly, by the second inequality in (3.23) Q is upper bounded by

(3.25) $$Q \leq \frac{1}{r} P[\chi^2_{\nu+2} > r\nu n_0] + P[\chi^2_{\nu} > r\nu n_0].$$

Combining (3.22), (3.24) and (3.25), we have

(3.26) $$EN = n_0 P[\chi^2_{\nu} \leq r\nu n_0] + \frac{1}{r} P[\chi^2_{\nu+2} > r\nu n_0] + \theta P[\chi^2_{\nu} > r\nu n_0]$$

for some $\theta \epsilon [0, 1]$.

The expected sample size EN given in (3.26) is a function of P^*, k, n_0 and λ, and it depends on P^* only through h_ν.

Lemma 3.2. For every P^*, k and first-stage sample size n_0,

(3.27) (1) $EN \geq 2\lambda^2 h_\nu^2$ for every λ,

(3.28) (2) $$\lim_{\lambda\to\infty} \frac{EN}{2\lambda^2 h_\nu^2} = 1.$$

Proof: By (3.7),

$$EN = E \max \{n_0, 2S_\nu^2 h_\nu^2/a^2\} \geq E\, 2S_\nu^2 h_\nu^2/a^2 = 2\lambda^2 h_\nu^2,$$

this proves (3.27).

The proof of (3.28) follows from (3.26).

Let N_0 be the sample required under the single-stage procedure for the LF configuration (given in (2.19)); the following theorem investigates the asymptotic relative efficiency of the two-stage procedure w.r.t. the single-stage procedure.

Theorem 3.3. For every P^*, k and first-stage sample size n_0,

(3.29) (1) $$\frac{EN}{N_0} \geq \left(\frac{h_\nu}{b}\right)^2$$ for every λ,

(3.30) (2) $$\lim_{\lambda\to\infty} \frac{EN}{N_0} = \left(\frac{h_\nu}{b}\right)^2.$$

Proof: The proof of this theorem follows from (2.22) and Lemma 3.2.

3.5 The Expected Misclassification Size.

The definition of misclassification about a population $\Pi_i (i = 1, 2, \ldots, k)$ was given in Chapter I and the expected misclassification size EM under the single-stage procedure was investigated in Lemma 2.4. In this section we give some results about EM under the two-stage procedure. It is easily seen that the supremum of EM over all mean vector $\vec{\mu}$ is achieved when μ_i is either $\mu_0 + \delta_1^*$ or $\mu_0 + \delta_2^*$ $(i = 1, 2, \ldots, k)$. In particular, it is achieved at $\vec{\mu}^0$ defined in (2.19). Hence we restrict our attention to $\vec{\mu}^0$ only and denote it by EM_0. The exact mathematical expression for EM_0 cannot be obtained due to the difficulty that the sample size N is a chance variable. However, both lower and upper bounds on EM_0 can be derived and the asymptotic behavior of EM_0 (as $\lambda \to \infty$) can be examined based on those bounds.

Lemma 3.3. For every λ, we have

$$(3.31) \qquad EM_0 \geq k[1 - \Phi(\sqrt{\tfrac{EN}{2}}/\lambda)].$$

Proof: It follows from the proof of Lemma 2.4 that

$$EM_0 = \sum_{n=n_0}^{\infty} k[1 - \Phi(\sqrt{\tfrac{n}{2}}/\lambda)]\cdot P[N = n] = Ek[1 - \Phi(\sqrt{\tfrac{N}{2}}/\lambda)]$$

where the expection is over N space. Consider $g(N) = 1 - \Phi(\sqrt{\frac{N}{2}}/\lambda)$ as a function of N. Since for $\lambda > 0$

$$\frac{d^2}{dN^2} g(N) = \frac{1}{8\lambda\sqrt{\pi N}} e^{-\frac{N}{4\lambda^2}} [\frac{1}{N} + \frac{1}{2\lambda^2}] > 0,$$

g(N) is a concave function of N. It follows from Jensen inequality that $Eg(N) \geq g(EN)$, which completes the proof.

To establish an upper bound on EM_0, we first state a result given in [13: p. 166].

Lemma 3.4. (Feller-Laplace). For every $z > 0$,

$$(3.32) \qquad \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}(\frac{1}{z} - \frac{1}{z^3}) < \int_z^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx < \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \frac{1}{z}.$$

When z is not too small, we can write the approximation

$$(3.33)\qquad \int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx \approx \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \cdot \frac{1}{z}.$$

Lemma 3.5. For every λ and the first-stage sample size n_0,

$$(3.34)\qquad EM_0 < k\{[1-\Phi(\sqrt{\tfrac{n_0}{2}}/\lambda)]P[\chi^2_\nu \le r\nu n_0] + \frac{\nu+h^2_\nu}{(\nu-1)h_\nu} f_\nu(h_\nu)P[\chi^2_{\nu-1} > r\nu n_0(1+\frac{h^2_\nu}{\nu})]\}$$

where $f_\nu(\cdot)$ is the density function of Student's t distribution with ν degrees of freedom and r is defined in (3.20).

Proof: By (3.21),

$$EM_0 = k\{[1-\Phi(\sqrt{\tfrac{n_0}{2}}/\lambda)]P[N=n_0] + \sum_{n=n_0+1}^{\infty}[1-\Phi(\sqrt{\tfrac{n}{2}}/\lambda)]P[N=n]\}$$

$$= k\{[1-\Phi(\sqrt{\tfrac{n_0}{2}}/\lambda)]P[\chi^2_\nu \le r\nu n_0] + \sum_{n=n_0+1}^{\infty}[1-\Phi(\sqrt{\tfrac{n}{2}}/\lambda)]P[r\nu(n-1)<\chi^2_\nu \le r\nu n]\}$$

$$\equiv k\{I_1 + I_2\};$$

we want to find an upper bound on I_2.

Using Lemma (3.4) and the first inequality in (3.23), we have

$$I_2 \le \sum_{n=n_0+1}^{\infty} \frac{\sqrt{2}\,\lambda}{\sqrt{2\pi n}} e^{-\frac{n}{4\lambda^2}} \int_{r\nu(n-1)}^{r\nu n} \frac{1}{2^{\nu/2}\Gamma(\frac{\nu}{2})} y^{\frac{\nu}{2}-1} e^{-\frac{y}{2}}\,dy$$

$$< \sum_{n=n_0+1}^{\infty} \sqrt{\frac{\nu}{\pi}}\,[2^{\frac{\nu}{2}+1}\Gamma(\tfrac{\nu}{2})h_\nu]^{-1} \int_{r\nu(n-1)}^{r\nu n} y^{\frac{\nu-3}{2}} e^{-\frac{1}{2}(1+\frac{h^2_\nu}{\nu})y}\,dy$$

$$= \sqrt{\frac{\nu}{\pi}}\,\frac{\Gamma(\frac{\nu-1}{2})}{2\Gamma(\frac{\nu}{2})h_\nu}\,[1+\frac{h^2_\nu}{\nu}]^{-(\frac{\nu-1}{2})} P[\chi^2_{\nu-1} > r\nu n_0(1+\frac{h^2_\nu}{\nu})]$$

$$= \frac{\nu+h^2_\nu}{(\nu-1)h_\nu} f_\nu(h_\nu)P[\chi^2_{\nu-1} > r\nu n_0(1+\frac{h^2_\nu}{\nu})],$$

which yields the desired result.

Theorem 3.4. For every P^*, k and first-stage sample size n_0,

$$(3.35)\qquad k[1-\Phi(h_\nu)] \le \lim_{\lambda\to\infty} EM_0 \le \frac{k(\nu+h^2_\nu)}{(\nu-1)h_\nu} f_\nu(h_\nu)$$

where $f_\nu(\cdot)$ is the density function of Student's t distribution with ν degrees of freedom.

Proof: The lower bound follows from Lemma 3.3 and (3.28), the upper bound follows from Lemma 3.5.

Remark 3.1. When ν is fairly large, the ratio $\frac{\nu+h_\nu^2}{\nu-1}$ is approximately 1, $f_\nu(h_\nu)$ is approximately $\varphi(h_\nu)$ where $\varphi(\cdot)$ is the standard normal density defined in (1.5). Applying (3.33), the upper bound in (3.35) is approximately $k[1 - \Phi(h_\nu)]$. Hence when ν is fairly large, $\lim_{\lambda\to\infty} EM_0$ is approximately $k[1 - \Phi(h_\nu)]$.

CHAPTER IV A Sequential Procedure

4.1 Introduction.

In Chapter III we have attacked the problem with unknown σ^2 by a two-stage sampling procedure based on the spirit of a procedure which was originally proposed by Stein [30] and introduced to the field of ranking and selection problems by Bechhofer, Dunnett and Sobel [4]. The asymptotic behavior of the expected sample size required to satisfy the probability requirement (1.4) for our problem and the expected misclassification size were examined there.

The expected sample size of Stein's procedure has been investigated in several papers; in particular, see [27], [28] and [30]. It has been shown that its relative efficiency, $\frac{N_0}{EN}$, is uniformly less than 1 (for all values of σ^2 and the first-stage sample size n_0), and that if n_0 "is poorly chosen, the ratio $\frac{N_0}{EN}$ may be quite small" [27: p. 2]. The relative inefficiency has been explained as being (at least partly) due to the fact that the information of the observations in the second stage is not utilized in estimating the unknown parameter σ^2. This gives us the idea of performing the experiment so that σ^2 can be estimated sequentially. A random stopping rule is developed by a few authors to solve the problem considered by Stein and the behavior of the expected sample size under this rule, especially when σ^2 is large, has been studied in [1], [7] and [29]. The same approach has been used recently in the problem of selecting the best of k normal populations in a paper by Robbins, Sobel and Starr [26].

In this chapter a sequential procedure which adopts the spirit of the above random stopping rule is developed for our problem to serve as an alternative to the two-stage procedure given in Chapter III. It provides an "approximate" solution to our problem in the sense that the PCD under this rule may be slightly less than P^* for some values of the unknown parameter σ^2. The asymptotic relative efficiency and the expected misclassification size under this rule will be investigated; some comparisons with that of the two-stage procedure will be made.

In recent years, some ranking and selection problems have been treated sequentially by several authors (e.g., see [5]), but our approach is different. It should be pointed out that our present approach is sequential mainly because we estimate the unknown parameter σ^2 sequentially, while the "usual" sequential ranking and selection problems is within the general framework of Wald's sequential decision theory and in general does not involve any nuisance parameter. Hence these two approaches have entirely different motivations.

4.2 The Proposed Rule and its Asymptotic Relative Efficiency.

When μ_0 is known, we propose the following rule:

Definition 4.1. Rule R_{3A}:

(1) We observe the sequence of random vectors $\vec{X}_j = \{X_{1j}, X_{2j}, \ldots, X_{kj}\}$ defined in (2.1), one vector at a time, stop with $\vec{X}_N$ where

(4.1) $$N \text{ is the first integer } n \geq 2 \text{ such that } S_\nu^2 \leq \frac{na^2}{w_\nu^2}, \quad a \text{ is defined}$$

in (1.9), $\nu = k(n-1)$, w_ν satisfies (3.1) and

$$S_\nu^2 = \frac{1}{\nu} \sum_{i=1}^{k} \sum_{j=1}^{n} [X_{ij} - \frac{1}{n}(\sum_{j=1}^{n} X_{ij})]^2.$$

(2) Let the observed N value in (1) be n. Compute

(4.2) $$\bar{X}_i = \frac{1}{n} \sum_{j=1}^{n} X_{ij} \quad \text{for } i = 1, 2, \ldots, k$$

and apply the decision rule defined in (2.5).

When μ_0 is unknown, the following rule is proposed:

Definition 4.2. Rule R_{3B}:

(1) We observe the sequence $\vec{X}_j = \{X_{0j}, X_{1j}, \ldots, X_{kj}\}$ defined in (2.3), one vector at a time, stop with $\vec{X}_N$ where

(4.3) $$N \text{ is the first integer } n \geq 2 \text{ such that } S_\nu^2 \leq \frac{na^2}{2h_\nu^2}, \quad a \text{ is defined}$$

in (1.9), $\nu = (k+1)(n-1)$, h_ν satisfies (3.12) with $P = P^*$ and

$$S_\nu^2 = \frac{1}{\nu} \sum_{i=0}^{k} \sum_{j=1}^{n} [X_{ij} - \frac{1}{n}(\sum_{j=1}^{n} X_{ij})]^2.$$

(2) Let the observed N value in (1) be n. Compute

(4.4) $$\bar{X}_i = \frac{1}{n} \sum_{j=1}^{n} X_{ij} \quad \text{for } i = 0, 1, \ldots, k$$

and apply the decision rule defined in (2.6).

Through the rest of this chapter we prove some results about the properties of the above proposed rule. We shall restrict our attention to R_{3B} only; it is easily seen that analogous arguments also apply to R_{3A}.

Lemma 4.1. For every $\vec{\mu}$ and every σ^2,

$$P[N < \infty | \vec{\mu}, \sigma^2; R_{3B}] = 1. \tag{4.5}$$

Proof: By the strong law of large numbers, $\lim_{\nu \to \infty} S_\nu^2 = \sigma^2$ a.s. Since by Theorem 3.2 $h_\nu \to b$, we have

$$P[N = \infty | \vec{\mu}, \sigma^2; R_{3B}] = P\{\bigcap_{n=2}^{\infty} [\frac{S_\nu^2}{n} > \frac{a^2}{2h_\nu^2}]\} = 0.$$

The following theorem states a relationship between the sample sizes required for the two-stage procedure and the sequential procedure. Let N_t and N_s denote the sample size required under R_{2B} (which is defined in Chapter III) and R_{3B} respectively. Then

Theorem 4.1. For every first-stage sample size n_0 in R_{2B}, we have

$$[N_t = n_0] \subset [N_s \leq N_t]. \tag{4.6}$$

Proof: Let $\mathfrak{X} = \{\omega: \omega = (\vec{x}_1, \vec{x}_2, \ldots)\}$ be the sample space. Since $\forall\ \omega \in \mathfrak{X}$ we have $N_t(\omega) \geq n_0$, it suffices to show that $[N_t = n_0] \subset [N_s \leq n_0]$.

Let $\{B_n\}$ and $\{C_n\}$ denote the terminal sets for R_{2B} and R_{3B} respectively; i.e.,

$$B_n = \{\omega: \omega \in \mathfrak{X}, N_t = n\} \quad \text{for} \quad n = n_0, n_0+1, \ldots,$$

$$C_n = \{\omega: \omega \in \mathfrak{X}, N_s = n\} \quad \text{for} \quad n = 2, 3, \ldots.$$

Then for $\nu = (k+1)(n_0-1)$, it follows from (3.7) that

$$\omega \in B_{n_0} \Leftrightarrow S_\nu^2(\omega) \leq \frac{n_0 a^2}{2h_\nu^2},$$

this implies that either there exists an $n' < n_0$ such that $\omega \in C_{n'}$ or $\omega \in C_{n_0}$. Hence $\omega \in \bigcup_{n=2}^{n_0} C_n$ or equivalently, $\omega \in [N_s \leq n_0]$.

Corollary. For $\nu = (k+1)(n_0-1)$,

$$P[N_s \leq N_t] \geq P[N_t = n_0] = P[\chi_\nu^2 \leq \frac{\nu n_0}{2\lambda^2 h_\nu^2}]. \tag{4.7}$$

In particular, $\lim_{\lambda\to 0} P[N_s \le N_t] = 1$ for every $n_0 \ge 2$.

As mentioned previously, the relative efficiency and the expected sample size of the random stopping rule, developed to provide an alternative solution to the one-population problem considered by Stein, have been studied by a number of authors. Special attention has been paid to the case when $\lambda \to \infty$ (either $\sigma^2 \to \infty$ or $a \to 0$ or both); and general results were given along somewhat different lines by Anscombe [1], [2], Chow and Robbins [7] and Gleser, Robbins and Starr [14]. For small and moderate values of λ, the relative efficiency of that procedure was investigated by Starr [29]. No exact mathematical expression of the expected sample size (EN) function was obtained. However, both Ray [24] and Robbins [25] have given a computing formula and some numerical results based on a slightly modified procedure, which terminates only at odd integers.

In the following we shall use some of those results to investigate the relative efficiency of the rule R_{3B}. We first observe that

Remark 4.1. For every $n \ge 2$ and for S_ν^2 given in (1) of Definition 4.2, $\frac{\nu S_\nu^2}{\sigma^2}$ is distributed as $V_1 + V_2 + \dots + V_{n-1}$ where the V's are independently identically distributed chi-square chance variables with $(k+1)$ degrees of freedom. (In fact, the V's can be obtained by using Helmert Transformation.)

Definition 4.3. We define a sequence of real numbers $\{q_j\}_1^\infty$ by

$$(4.8) \qquad q_j = \left(\frac{k+1}{2\lambda^2}\right)\left[\frac{j(j+1)}{h^2_{(k+1)j}} - \frac{(j-1)j}{h^2_{(k+1)(j-1)}}\right] \quad \text{for } j = 1, 2, \dots$$

where $h_0 = 0$ and $h_{(k+1)j}$ satisfies (3.12) with $P = P^*$ for $j \ge 1$.

The following theorem gives the bounds on the c.d.f. of the random sample size N under the rule R_{3B}.

Theorem 4.2. For every fixed $n \ge 2$,

$$(4.9) \qquad \chi^2_\nu\left(\frac{\nu n}{2\lambda^2 h^2_\nu}\right) \le P[N \le n] \le 1 - \prod_{j=1}^{n-1} [1 - \chi^2_{(k+1)}(q_j)]$$

where $\nu = (k+1)(n-1)$.

Proof: It follows from (4.3) that

$$[N > n] = \bigcap_{j=2}^{n} \left[\frac{(k+1)(j-1)}{\sigma^2} S^2_{(k+1)(j-1)} > \frac{(k+1)(j-1)j}{2\lambda^2 h^2_\nu}\right]$$

$$= [V_1 > q_1, \sum_{j=1}^{2} V_j > \sum_{j=1}^{2} q_j, \ldots, \sum_{j=1}^{n-1} V_j > \sum_{j=1}^{n-1} q_j].$$

Since

$$[\sum_{j=1}^{n-1} V_j > \sum_{j=1}^{n-1} q_j] \supset [V_1 > q_1, \sum_{j=1}^{2} V_j > \sum_{j=1}^{2} q_j, \ldots, \sum_{j=1}^{n-1} V_j > \sum_{j=1}^{n-1} q_j] \supset \bigcap_{j=1}^{n-1} [V_j > q_j],$$

it follows that

(4.10) $$1 - \chi^2_\nu\left(\frac{\nu n}{2\lambda^2 h^2_\nu}\right) \geq P[N > n] \geq \prod_{j=1}^{n-1} [1 - \chi^2_{(k+1)}(q_j)],$$

and the theorem is proved by taking complements.

Corollary: For every fixed $n \geq 2$,

(4.11) (1) $\lim_{\lambda \to 0} P[N \leq n] = 1,$

(4.12) (2) $\lim_{\lambda \to \infty} P[N \leq n] = 0.$

In particular, the c.d.f. of N converges to a degenerate distribution as $\lambda \to 0$; i.e.,

(4.13) $\lim_{\lambda \to 0} P[N = 2] = 1.$

Remark 4.2. (4.12) implies that as $\lambda \to \infty$, $N \to \infty$ in probability which is also a consequence of Chow and Robbins [7], see (4.15) in Lemma 4.2 below.

For large values of λ, we first state two lemmas which are due to Chow and Robbins [7]:

Lemma 4.2. (Chow and Robbins). Let y_n $(n = 1, 2, \ldots)$ be any sequence of random variables such that $y_n > 0$ a.s., $\lim_{n\to\infty} y_n = 1$ a.s., let $f(n)$ be any sequence of constants such that $f(n) > 0$, $\lim_{n\to\infty} f(n) = \infty$, $\lim_{n\to\infty} f(n)/f(n-1) = 1$, and for each $t > 0$, define

(4.14) $N = N(t) =$ smallest $n \geq 1$ such that $y_n \leq f(n)/t.$

Then N is well defined and nondecreasing as a function of t, (4.15) $\lim_{t\to\infty} N = \infty$ a.s., $\lim_{t\to\infty} EN = \infty$

and

(4.16) $\lim_{t\to\infty} f(N)/t = 1$ a.s.

Lemma 4.3. (Chow and Robbins). If the conditions of Lemma 4.2 hold and if also $E(\sup_n y_n) < \infty$, then

(4.17) $\lim_{t\to\infty} Ef(N)/t = 1.$

Let N_0 be the sample size required under the single-stage procedure for the LF configuration (given in (2.19)); the following theorem investigates the asymptotic relative efficiency of the sequential proceudre w.r.t. the single-stage procedure.

Theorem 4.3. Let N_0 be defined in (2.22) and N be the random sample size defined in (4.3). Then

(4.18) (1) $\lim_{\lambda\to\infty} \frac{N}{N_0} = 1$ a.s.,

(4.19) (2) $\lim_{\lambda\to\infty} \frac{EN}{N_0} = 1.$

Using the terminology in [7], it follows from (4.19) that the rule R_{3B} is asymptotically relatively efficient.

Proof: For $\nu = (k+1)(n-1)$, set $y_n = \frac{S_\nu^2}{\sigma^2}$, $f(n) = n(\frac{b}{h_\nu})^2$ and $t = 2\lambda^2 b^2$. Since y_n is distributed as $\frac{1}{(k+1)(n-1)}(V_1 + V_2 + \dots + V_{n-1})$ where the V's are i.i.d. chi-square chance variables each with $(k+1)$ degrees of freedom, it follows then from the strong law of large numbers that $\lim_{n\to\infty} y_n = 1$ a.s. the rest of the conditions in Lemma 4.2 are easily seen to be satisfied since (by Theorem 3.2) $h_\nu \to b$. Hence (4.18) is proved.

To prove (4.19), by Lemma 4.3 it suffices to show that $E(\sup_n y_n) < \infty$. Let $c > 1$ be any real number. Then

$$P[\sup_n y_n > c] = P\{\bigcup_{n=1}^{\infty} [\frac{1}{(k+1)n} \sum_{j=1}^{n} V_j > c]\}$$

$$\leq \sum_{n=1}^{\infty} P[\sum_{j=1}^{n} V_j > (k+1)nc] \leq \sum_{n=1}^{\infty} P[|\sum_{j=1}^{n} V_j - (k+1)n| > (k+1)n(c-1)]$$

$$\leq \sum_{n=1}^{\infty} \frac{E[\sum_{j=1}^{n} V_j - (k+1)n]^4}{[(k+1)n]^4(c-1)^4}$$

where the last inequality follows from Markov inequality. Now for every fixed n, $\sum_{j=1}^{n} V_j$ is a chi-square chance variable with $(k+1)n$ degrees of freedom. By elementary calculations it is easily seen that the fourth central moment of a chi-square chance variable with d degrees of freedom is $12d(d+4)$. Hence

$$E[\sum_{j=1}^{n} V_j - (k+1)n]^4 = 12(k+1)n[(k+1)n + 4] \leq 60(k+1)^2n^2$$

which implies that for every $c > 1$,

$$P[\sup_n y_n > c] \leq \frac{60}{(k+1)^2(c-1)^4} \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{M}{(c-1)^4} \tag{4.20}$$

for some finite number M that does not depend on c. Thus

$$E(\sup_n y_n) \leq 2 + \sum_{j=1}^{\infty} (2+j)P[2+j-1 < \sup_n y_n \leq 2+j]$$

$$\leq 2 + \sum_{j=1}^{\infty} (2+j)P[\sup_n y_n > 2+j-1]$$

$$\leq 2 + \sum_{j=1}^{\infty} \frac{M(2+j)}{j^4} \leq 2 + 3M \sum_{j=1}^{\infty} \frac{1}{j^3} < \infty$$

which completes the proof of (4.19).

4.3 The PCD Function and its Asymptotic Behavior.

As mentioned at the beginning of this chapter, the proposed rule provides only an "approximate" solution to our problem in the sense that the PCD under this rule may be slightly less than P^* for some values of the unknown parameter σ^2, or equivalently, λ. In this section we examine PCD and its asymptotic behavior as a function of λ.

For the covariance matrix Σ specified in (2.20), we first define a function

$$\beta(\sqrt{\tfrac{n}{2}}/\lambda) = \int_{-\infty}^{\sqrt{\frac{n}{2}}/\lambda}\int_{-\infty}^{\sqrt{\frac{n}{2}}/\lambda}\cdots\int_{-\infty}^{\sqrt{\frac{n}{2}}/\lambda} \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}y'\Sigma^{-1}y} \prod_{i=1}^{k} dy_i \tag{4.21}$$

(note that β depends on n and λ only through the ratio $\sqrt{\frac{n}{2}}/\lambda$).

Then it is easily seen for any mean vector $\vec{\mu}$, the conditional PCD given $N = n$ is lower bounded by

$$P[CD|\vec{\mu}, \lambda, N=n] \geq P[CD|\vec{\mu}^0, \lambda, N=n] = \beta(\sqrt{\tfrac{n}{2}}/\lambda)$$

for every n, where $\vec{\mu}^0$ is given in (2.19). Hence we have

$$P[CD|\vec{\mu}, \lambda, R_{3B}] \geq P[CD|\vec{\mu}^0, \lambda; R_{3B}] = E\beta(\sqrt{\tfrac{N}{2}}/\lambda) \tag{4.22}$$

where the expectation is taken overn N space (it should be observed from (4.3) that the distribution of N here depends on the parameter λ).

<u>Theorem 4.4</u>. For every mean vector $\vec{\mu}$,

$$(1)\quad \lim_{\lambda\to 0} P[CD|\vec{\mu}, \lambda; R_{3B}] = 1, \tag{4.23}$$

$$(2)\quad \lim_{\lambda\to \infty} P[CD|\vec{\mu}, \lambda; R_{3B}] \geq P^*. \tag{4.24}$$

<u>Proof</u>: By (4.22), we can restrict our attention to $\vec{\mu} = \vec{\mu}^0$ and work on $E\beta(\sqrt{\frac{N}{2}}/\lambda)$.

Since β is continuous and monotone increasing and $N \geq 2$ a.s., it follows that $\beta(\sqrt{\frac{N}{2}}/\lambda) \geq \beta(\frac{1}{\lambda})$ a.s. and

$$\lim_{\lambda\to 0} E\beta(\sqrt{\tfrac{N}{2}}/\lambda) \geq \lim_{\lambda\to 0} E\beta(\tfrac{1}{\lambda}) = \lim_{\lambda\to 0} \beta(\tfrac{1}{\lambda}) = 1. \tag{4.25}$$

This proves (4.23).

To prove (4.24), let $\{\lambda_j\}_1^\infty$ be an arbitrary but fixed monotone increasing sequence such that $\lim_{j\to\infty} \lambda_j = \infty$. By (4.18), $\lim_{j\to\infty} \frac{N}{2\lambda_j^2 b^2} = 1$ a.s. where b is such that $\beta(b) = P^*$. Since a.s. convergence is preserved by continuous mapping, it then follows

$$\lim_{j\to\infty} \beta(\sqrt{\tfrac{N}{2}}/\lambda_j) = \beta(b) \quad \text{a.s.} \tag{4.26}$$

Let $F_j(\cdot)$ $(i = 1, 2, \ldots)$ and $F(\cdot)$ be the c.d.f. of $\beta(\sqrt{\frac{N}{2}}/\lambda_j)$ $(j = 1, 2, \ldots)$ and $\beta(b)$ respectively. Then it follows from (4.26) that

$$F_j(\cdot) \xrightarrow{d} F(\cdot). \tag{4.27}$$

But the sequence of chance variables $\{\beta(\sqrt{\frac{N}{2}}/\lambda_j)\}$ is uniformly bounded by 0 and 1 and $\beta(b)$ is a degenerate variable, so we must have $F_j(0) = F(0) = 0$

and $F_j(1) = F(1) = 1$ $(j = 1, 2, \ldots)$, which implies $F_j(0) \to F(0)$ and $F_j(1) \to F(1)$. Applying Helly-Bray Lemma (see, e.g., [21: p. 180]),

$$\lim_{\lambda \to \infty} E\beta\left(\sqrt{\frac{N}{2}}/\lambda_j\right) = \lim_{j \to \infty} \int_0^1 y\,dF_j(y) = \int_0^1 y\,dF(y) = P^*.$$

Since the sequence $\{\lambda_j\}$ is arbitrarily chosen, the proof of (4.24) is completed.

Theorem 4.4 establishes the properties of PCD as λ takes the extreme values. The theoretical part of the problem concerning the behavior of PCD for moderate values of λ is still unsolved. Starr [29] has obtained some numerical results about the sequential confidence interval for the mean in the one-population problem. His results show that the confidence coefficient given by his sequential procedure is slightly below the fixed level for moderate values of λ. We do expect a similar situation under the rule R_{3B} proposed in Section 4.2 of this chapter. However, the author has not reached any theoretical or numerical conclusions.

4.4 The Expected Misclassification Size.

In this section we examine the expected misclassification size EM under R_{3B}. It is easily seen that the supremum of EM over all mean vector $\vec{\mu}$ is achieved when μ_i is either $\mu_0 + \delta_1^*$ or $\mu_0 + \delta_2^*$ $(i = 1, 2, \ldots, k)$. In particular, it is achieved at $\vec{\mu}^0$ defined in (2.19). Hence we restrict our attention to $\vec{\mu}^0$ only and denote it by EM_0. We first give a lower bound on EM_0 for all λ.

Lemma 4.4: For every $\lambda > 0$,

(4.28) $$EM_0 \geq k\left[1 - \Phi\left(\sqrt{\frac{EN}{2}}/\lambda\right)\right].$$

Proof: The proof of this lemma is similar to that of Lemma 3.3.

The following theorem shows that for extreme values of λ, EM_0 under the sequential procedure is the same as that under the single-stage procedure.

Theorem 4.5. For b defined in (2.21) with $P = P^*$,

(4.29) (1) $\lim_{\lambda \to 0} EM_0 = 0,$

(4.30) (2) $\lim_{\lambda \to \infty} EM_0 = k[1 - \Phi(b)].$

Proof: The proof of this theorem is similar to that of Theorem 4.4 with $\beta(\sqrt{\frac{N}{2}}/\lambda) = k[1 - \Phi(\sqrt{\frac{N}{2}}/\lambda)]$.

APPENDIX A

A Property of Certain Multivariate Normal Distributions

For arbitrary but fixed real numbers $\rho \epsilon (0, 1)$, $c \epsilon (-\infty, \infty)$, $\sigma^2 > 0$ and positive integer k, we define (for every fixed $r = 0, 1, \dots, k$) the $(k \times k)$ covariance matrix $\Sigma_r = (\sigma_{ij})$ by

$$(A.1) \qquad \sigma_{ij} = \begin{cases} \sigma^2 & \text{if } i = j, \\ \rho\sigma^2 & \text{if } i, j \epsilon \{1, 2, \dots, r\} \text{ or } i, j \epsilon \{r+1, r+2, \dots, k\}, \\ -\rho\sigma^2 & \text{if } i \epsilon \{1, 2, \dots, r\} \text{ and } j \epsilon \{r+1, r+2, \dots, k\}; \end{cases}$$

i.e., Σ_r has the following structure:

$$(A.2) \qquad \Sigma_r = \sigma^2 \begin{array}{cl} \begin{array}{cccccc} 1\ 2 & \cdots & r & r+1 & \cdots & k \end{array} & \\ \left(\begin{array}{ccc|ccc} 1 & & \rho & & & \\ & \ddots & & & -\rho & \\ \rho & & 1 & & & \\ \hline & & & 1 & & \rho \\ & -\rho & & & \ddots & \\ & & & \rho & & 1 \end{array}\right) & \begin{array}{l} 1 \\ 2 \\ \vdots \\ r \\ r+1 \\ \vdots \\ k \end{array} \end{array} .$$

We also define the multivariate normal probability integral $P(r)$ by

$$(A.3) \qquad P(r) = P_{\rho,c}(r) = \int_{-\infty}^{c} \int_{-\infty}^{c} \cdots \int_{-\infty}^{c} \frac{1}{(2\pi)^{k/2} |\Sigma_r|^{1/2}} e^{-\frac{1}{2} \underset{\sim}{y}' \Sigma_r^{-1} \underset{\sim}{y}} \prod_{i=1}^{k} dy_i .$$

It should be observed that for either $r = 0$ or $r = k$, $P(r)$ is the probability integral of an equicorrelated multivariate normal distribution. Let the integer $[x]$ denote the largest integer $\leq x$; we say $P(r)$ is strictly decreasing at r ($P(r)\downarrow r$) if $P(r+1) - P(r) < 0$, and we say $P(r)$ is strictly increasing at r ($P(r)\uparrow r$) if $P(r+1) - P(r) > 0$. The purpose of this section is to prove the following

<u>Theorem A.1</u>. For every $\rho \epsilon (0, 1)$, $c \epsilon (-\infty, \infty)$ and $\sigma^2 > 0$, we have

(A.4) (1) $P(r) = P(k-r)$ for $r = 0, 1, \dots, k$;

(A.5) (2) $P(r)\downarrow r$ for $r = 0, 1, \dots, [\frac{(k-2)}{2}]$;

(A.6) (3) $P(r)\uparrow r$ for $r = k-1, k-2, \dots, [\frac{(k+1)}{2}]$.

[We shall refer to (A.4) as the symmetry property of $P(r)$ and (A.5) and (A.6) as the monotonicity properties of $P(r)$.]

Corollary: $P(r)$ achieves a unique minimum at $r = \frac{k}{2}$ when k is even and a common minimum at $r = \frac{(k-1)}{2}$ and $r = \frac{(k+1)}{2}$ when k is odd.

Before we prove this theorem we first prove a lemma dealing with symmetric distributions. Let $f(z)$ be a density function and $F(z)$ be its corresponding c.d.f. For arbitrary but fixed real numbers $\eta \epsilon (0, \infty)$, $s \epsilon (-\infty, \infty)$ and any positive integer k, we define for $r = 0, 1, \ldots, k$

$$\text{(A.7)} \qquad \beta(r) = \int_{-\infty}^{\infty} F^{r}(\eta z + s) F^{k-r}(-\eta z + s) f(z) dz,$$

and its first difference

$$\text{(A.8)} \qquad \Delta\beta(r) = \beta(r+1) - \beta(r), \quad r = 0, 1, \ldots, k-1.$$

Lemma A.1. If $f(z) = f(-z)$, then for every $\eta \epsilon (0, \infty)$ and $s \epsilon (-\infty, \infty)$,

$$\text{(A.9)} \qquad (1) \quad \beta(r) = \beta(k-r) \quad \text{for } r = 0, 1, \ldots, k;$$

$$\text{(A.10)} \qquad (2) \quad \Delta\beta(r) \leq 0 \quad \text{for } r = 0, 1, \ldots, [\tfrac{(k-2)}{2}];$$

$$\text{(A.11)} \qquad (3) \quad \Delta\beta(r) \geq 0 \quad \text{for } r = k-1, k-2, \ldots, [\tfrac{(k+1)}{2}].$$

Proof: Property (A.9) follows immediately by setting $u = -z$ in the integral on the r.h.s. of (A.7).

To prove (A.10), we first define a function

$$\text{(A.12)} \qquad H(z) = [F(\eta z + s) - F(-\eta z + s)] f(z) \quad \text{for } z \epsilon (-\infty, \infty).$$

Then it is easily seen that

$$\text{(A.13)} \qquad H(z) = -H(-z) \quad \text{for } z \epsilon (-\infty, \infty)$$

and since $\eta > 0$

$$\text{(A.14)} \qquad H(z) \geq 0 \quad \text{for } z \epsilon (0, \infty).$$

For every fixed $r \leq \frac{(k-1)}{2}$,

$$\text{(A.15)} \qquad \Delta\beta(r) = \int_{-\infty}^{\infty} F^{r}(\eta z + s) F^{k-r-1}(-\eta z + s) [F(\eta z + s) - F(-\eta z + s)] f(z) dz$$

$$= \int_0^\infty F^r(\eta z + s)F^{k-r-1}(-\eta z + s)H(z)dz + \int_{-\infty}^0 F^r(\eta z + s)F^{k-r-1}(-\eta z + s)H(z)dz.$$

Consider the second integral I_2 on the r.h.s. of the above expression. Setting $u = -z$ and applying (A.13), we have $I_2 = -\int_0^\infty F^{k-r-1}(\eta u + s)F^r(-\eta u + s)H(u)du$. Substituting this in (A.15) gives

$$\text{(A.16)} \qquad \Delta\beta(r) = \int_0^\infty F^r(\eta z + s)F^r(-\eta z + s)H(z)[F^{k-2r-1}(-\eta z + s) - F^{k-2r-1}(\eta z + s)]dz.$$

Since by (A.14) $F^r(\eta z + s)F^r(-\eta z + s)H(z) \geq 0$ for $z \in (0, \infty)$ and $F(-\eta z + s) \leq F(\eta z + s)$, it follows that $\Delta\beta(r) \leq 0$ for $k - 2r - 1 > 0$ or equivalently, for $r \leq [\frac{(k-2)}{2}]$; and $\Delta\beta(r) = 0$ for $r = \frac{(k-1)}{2}$. This proves (A.10).

The result (A.11) is an immediate consequence of (A.9) and (A.10).

Remark A.1. If the c.d.f. $F(z)$ is strictly increasing in $(-\infty, \infty)$, then every inequality in the proof of the above lemma will be a strict inequality and hence the inequalities in (A.10) and (A.11) are strict inequalities.

Proof of the theorem: Without loss of generality, we assume $\sigma^2 = 1$.

Let $Z_0, Z_1, \ldots, Z_k$ be independent standard normal chance variables, $\varphi(\cdot)$ and $\Phi(\cdot)$ be the density function and c.d.f., respectively, of the standard normal distribution. For arbitrary but fixed $\rho \in (0, 1)$ and $c \in (-\infty, \infty)$, let $\eta > 0$ satisfy

$$\text{(A.17)} \qquad \frac{\eta^2}{\eta^2 + 1} = \rho > 0$$

and let

$$\text{(A.18)} \qquad s = c\sqrt{\eta^2 + 1}.$$

For fixed $r (0 \leq r \leq k)$, we define

$$\text{(A.19)} \qquad Y_i = \begin{cases} (Z_i - \eta Z_0)/\sqrt{\eta^2 + 1} & \text{for } i = 1, 2, \ldots, r; \\ (\eta Z_0 - Z_i)/\sqrt{\eta^2 + 1} & \text{for } i = r+1, r+2, \ldots, k. \end{cases}$$

Then $(Y_1, Y_2, \ldots, Y_k)$ follows a multivariate normal distribution with mean vector 0 and covariance matrix Σ_r defined in (A.2) with $\sigma^2 = 1$. Hence

$$(A.20) \qquad = P[Z_i \le \eta Z_0 + s,\ Z_j > \eta Z_0 - s\ (\begin{smallmatrix} i = 1, 2, \ldots, r \\ j = r+1, r+2, \ldots, k \end{smallmatrix})]$$

$$= \int_{-\infty}^{\infty} \Phi^r(\eta z + s)\Phi^{k-r}(-\eta z + s)\varphi(z)dz.$$

The rest of the argument follows from Lemma A.1. This completes the proof.

Corollaries. (1) Let $(U_1, U_2, \ldots, U_k)$ have the joint distribution $N(0, \Sigma_r)$ and let $(V_1, V_2, \ldots, V_k)$ have the joint distribution $N(0, \Sigma_s)$. Let $U = \max_{1 \le i \le k} U_i$, $V = \max_{1 \le i \le k} V_i$. If $|r - \frac{k}{2}| < |s - \frac{k}{2}|$, then U is stochastically larger than V.

(2) Let $(U_1, U_2, \ldots, U_k)$ and $(V_1, V_2, \ldots, V_k)$ follow multivariate t distributions; and the associated correlation matrices be Σ_r and Σ_s respectively. Let $U = \max_{1 \le i \le k} U_i$, $V = \max_{1 \le i \le k} V_i$. If $|r - \frac{k}{2}| < |s - \frac{k}{2}|$, then U is stochastically larger than V.

Example. We consider the special case $k = 2$ and $c = 0$. It is well-known that if (U_1, U_2) follows a bivariate normal distibution with means 0, a common but arbitrary variance σ^2 and correlation coefficient ρ, then $g(\rho) = P[U_1 \le 0, U_2 \le 0] = \frac{1}{4} + \frac{1}{2\pi}$ arc sin ρ (see, e.g., [8: p. 290]). If $\rho > 0$, then $g(\rho) > g(-\rho)$. Our result agrees with this statement because $g(\rho)$ corresponds to $P(r)$ for either $r = 0$ or $r = 2$ and $g(-\rho)$ corresponds to $P(r)$ for $r = 1$.

APPENDIX B

Convergence of the Percentage Points of a Sequence of Distributions

The purpose of this section is to prove the following theorem which deals with the relationship between convergence in distribution and convergence in their corresponding percentage points.

Theorem B.1: Let $\{F_n(\cdot)\}$ be a sequence of c.d.f.'s such that

(B.1) $F_n(\cdot) \xrightarrow{d} F(\cdot)$

and for arbitrary but fixed $P \epsilon (0, 1)$, let

(B.2) $\xi_n = \xi_n(P) = \inf\{x: F_n(x-0) \leq P \leq F_n(x)\}, \quad n = 1, 2, \ldots$

(B.3) $\xi = \xi(P) = \inf\{x: F(x-0) \leq P \leq F(x)\}.$

If $F(\cdot)$ is continuous and $F(s) > F(\xi)$ for every $s > \xi$, then

(B.4) $\xi_n \to \xi$ as $n \to \infty$.

Before we begin the proof of the theorem we first establish two lemmas.

Lemma B.1. A continuous c.d.f. $F(\cdot)$ is uniformly continuous in $(-\infty, \infty)$.

Proof: For every $\epsilon > 0$, let M be large enough such that $F(-M) < \frac{\epsilon}{2}$ and $F(M) > 1 - \frac{\epsilon}{2}$. Since $F(\cdot)$ is uniformly continuous in the close interval $[-M, M]$, there exists an δ such that $|x - x'| < \delta$ implies $|F(x) - F(x')| < \epsilon$ for every $x, x' \epsilon [-M, M]$. Since $F(\cdot)$ is monotone increasing, this shows that $|x - x'| < \delta$ implies $|F(x) - F(x')| < \epsilon$ for every $x, x' \epsilon (-\infty, \infty)$.

Lemma B.2. If a sequence of c.d.f.'s $F_n(\cdot) \xrightarrow{d} F(\cdot)$ and $F(\cdot)$ is continuous, then the convergence is uniform.

Proof: For every $\epsilon > 0$, let M be large enough such that $F(-M) < \frac{\epsilon}{2}$ and $F(M) > 1 - \frac{\epsilon}{2}$. Let δ be small enough such that $|x - x'| < \delta$ implies $|F(x) - F(x')| < \frac{\epsilon}{5}$ for every $x, x' \epsilon (-\infty, \infty)$. For the finite partition $x_0 < x_1 < x_2 < \ldots < x_R$ such that $x_i = i\delta - M$ $(i = 0, 1, \ldots, R)$ and $x_R \geq M$, let N_i be large enough such that for every $n \geq N_i$, $|F_n(x_i) - F(x_i)| < \frac{\epsilon}{5}$ $(i = 0, 1, \ldots, R)$. Let $N = \max_{0 \leq i \leq R} N_i$. Then for every $n > N$, if $x \epsilon [-M, M]$, there exists an integer i such that $x_i \leq x < x_{i+1}$ and

$$|F_n(x) - F(x)| \leq |F_n(x) - F_n(x_i)| + |F_n(x_i) - F(x_i)| + |F(x_i) - F(x)|$$

$$\leq |F_n(x_{i+1}) - F_n(x_i)| + |F_n(x_i) - F(x_i)| + |F(x_{i+1}) - F(x_i)|$$

$$\leq |F_n(x_{i+1}) - F(x_{i+1})| + |F(x_{i+1}) - F(x_i)| + |F(x_i) - F_n(x_i)| + |F_n(x_i)-F(x_i)|$$

$$+ |F(x_{i+1}) - F(x_i)| < \epsilon;$$

if $x < -M$, by the monotonicity property of distribution functions,

$$|F_n(x) - F(x)| \leq F_n(-M) + F(-M) \leq 2\epsilon;$$

a similar argument also holds for the case $x > M$. Hence for every $n > N$, $|F_n(x) - F(x)| < 2\epsilon$ for every $x \epsilon (-\infty, \infty)$. This completes the proof.

Proof of the theorem: Since $F(\cdot)$ is continuous, it follows by (A.3) that $F(\xi) = P$. We first prove $F_n(\xi_n) \to F(\xi) = P$. By (B.2), we always have $F_n(\xi_n) \geq P$ and, if a_n is any real number $\leq \xi_n$, then $F_n(a_n) < P$ $(n = 1, 2, \ldots)$. Suppose $F_n(\xi_n) \nrightarrow P$, then for some $\epsilon > 0$ there exists an infinite set I of positive integers such that $F_n(\xi_n) \geq P + \epsilon$ for every $n \epsilon I$. Let c be such that

(B.5) $$F(c) = P + \frac{\epsilon}{2} .$$

Then for every $n \epsilon I$ we have either $F_n(c) < P$ (if $c < \xi_n$) or $F_n(c) \geq P + \epsilon$ (if $c \geq \xi_n$). Thus $F_n(c) \nrightarrow F(c)$ and we reach a contradiction. Hence $F_n(\xi_n) \to F(\xi) = P$.

Since $F_n(\cdot) \to F(\cdot)$ uniformly, there exists an N such that $n > N$ implies $|F_n(x) - F(x)| < \frac{\epsilon}{2}$ for every $x \epsilon (-\infty, \infty)$, hence

(B.6) $$F_n(c) - P = (F_n(c) - F(c)) + (F(c) - P) \geq -\frac{\epsilon}{2} + \frac{\epsilon}{2} = 0.$$

This shows for every $n > N$, $\xi_n \leq c$. The lower bound of $\{\xi_n\}$ can be obtained in a similar manner. Hence the sequence $\{\xi_n\}$ is uniformly bounded. By Weierstrass Theorem, the sequence $\{\xi_n\}$ has at least one limit point.

Let s be any limit point of $\{\xi_n\}$, or equivalently, let $\xi_{n_j} \to s$ as $j \to \infty$. Then

(B.7) $$\Delta_j = |F_{n_j}(\xi_{n_j}) - F(s)| \leq |F_{n_j}(\xi_{n_j}) - F(\xi_{n_j})| + |F(\xi_{n_j}) - F(s)|.$$

By the uniform convergence of $\{F_n(\cdot)\}$ and the uniform continuity of $F(\cdot)$, we have $\Delta_j \to 0$. Hence

(B.8) $$F_{n_j}(\xi_{n_j}) \to F(s) \quad \text{as} \quad j \to \infty$$

But $F_{n_j}(\xi_{n_j}) \to F(\xi)$, hence we must have $F(s) = F(\xi)$, thus (by (B.3)) $s \geq \xi$. Since by the condition on F $s > \xi$ implies $F(s) > F(\xi)$, it then follows that $s = \xi$.

This shows that any subsequence of $\{\xi_n\}$ converges to ξ. Hence $\xi_n \to \xi$ as $n \to \infty$ and the proof is completed.

Corollaries: (1) Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of chance variables such that the Central Limit Theorem applies. Let $\xi_n(P)$ be the percentage point of the distribution of the chance variable $\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} EX_i\right) / \sqrt{\sum_{i=1}^{n} \text{Var } X_i}$ and $\xi(P)$ be the percentage point of the standard normal distribution. Then $\xi_n(P) \to \xi(P)$ for every $P \in (0, 1)$.

(2) Let $F(\cdot)$ be a c.d.f. which is continuous and strictly increasing, and $F_n(\cdot)$ be the sample c.d.f. of n independent observations from $F(\cdot)$. Let $\xi_n(P)$ and $\xi(P)$ be the corresponding percentage points of $F_n(\cdot)$ and $F(\cdot)$ respectively, as defined in (B.2) and (B.3) (note that $\{\xi_n(P)\}$ is a sequence of chance variables). Then $\xi_n(P) \to \xi(P)$ a.s. for every $P \in (0, 1)$.

The proof of this corollary follows from Theorem B.1 and Glivenko - Cantelli Theorem which states that $F_n(\cdot) \to F(\cdot)$ uniformly with probability 1 (see, e.g., [21: p. 20]). In application, we regard $\{\xi_n(P)\}$ as a sequence of estimators of $\xi(P)$ and it follows that the estimator of the percentage point of $F(\cdot)$ based on the percentage point of the sample c.d.f. is consistant.

Example: The following is an example in which the conditions in Theorem B.1 are not satisfied and the conclusion does not follow.

Let the sequence of c.d.f.'s $\{F_n(\cdot)\}$ be such that $F_n(x) \equiv 0$ for $x < 0$, $F_n(x) \equiv 1$ for $x > 1$ and

(B.9) $$F_n(x) = \begin{cases} \frac{1}{2} - (\frac{1}{2} - x)^n & \text{for } 0 \leq x < \frac{1}{2}, \\ x & \text{for } \frac{1}{2} \leq x \leq 1; \end{cases}$$

then the limiting distribution function is such that $F(x) = 0$ for $x < 0$, $F(x) = 1$ for $x > 1$ and

$$\text{(B.10)} \qquad F(x) = \begin{cases} \frac{1}{2} & \text{for } 0 \leq x < \frac{1}{2}, \\ x & \text{for } \frac{1}{2} \leq x \leq 1. \end{cases}$$

For $P = \frac{1}{2}$, we have $\xi_n = \frac{1}{2}$ $(n = 1, 2, \ldots)$, hence $\xi_n \nrightarrow \xi = 0$.

Table 1

Equicoordinate percentage points b of a multivariate normal distribution with mean vector 0 and covariance matrix Σ

K	P = 0.50	P = 0.75	P = 0.90	P = 0.95	P = 0.975	P = 0.99
1	0.0000000	0.6744898	1.2815516	1.6448537	1.9599640	2.3263479
2	0.6423429	1.1462928	1.6445631	1.9599246	2.2413975	2.5758290
3	0.8370415	1.3192980	1.8003977	2.1057358	2.3786364	2.7033911
4	0.9938965	1.4528031	1.9162111	2.2121205	2.4775016	2.7942727
5	1.0890009	1.5389483	1.9950311	2.2865328	2.5480781	2.8604419
6	1.1742510	1.6140189	2.0620112	2.3489679	2.6067571	2.9149993
7	1.2356655	1.6702228	2.1138621	2.3981570	2.6536084	2.9591380
8	1.2928724	1.7214957	2.1602823	2.4417695	2.6948543	2.9977379
9	1.3376440	1.7627532	2.1985565	2.4781993	2.7296540	3.0306329
10	1.3801626	1.8012938	2.2337781	2.5114623	2.7612442	3.0603289
12	1.4486915	1.8643484	2.2921627	2.5669962	2.8142841	3.1104789
14	1.5048107	1.9162434	2.3404131	2.6130015	2.8583140	3.1522062
16	1.5521539	1.9601991	2.3814180	2.6521756	2.8958693	3.1878656
18	1.5929848	1.9982343	2.4169983	2.6862232	2.9285565	3.2189529
20	1.6288041	2.0316945	2.4483726	2.7162884	2.9574557	3.2464759

Note: The table value $b = b(P,k)$ is the solution of the equation

$$P = \int_{-\infty}^{b}\int_{-\infty}^{b}\cdots\int_{-\infty}^{b} \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}}\, e^{-\frac{1}{2}y'\Sigma^{-1}y} \prod_{i=1}^{k} dy_i$$

where the convariance matrix Σ is $(k \times k)$ and has the structure

$$\begin{array}{cc} \begin{array}{cc} 1\;2\ldots m & m+1\ldots k \end{array} & \\ \Sigma = \left(\begin{array}{ccc|ccc} 1 & & \frac{1}{2} & & & \\ & \ddots & & & -\frac{1}{2} & \\ \frac{1}{2} & & 1 & & & \\ \hline & & & 1 & & \frac{1}{2} \\ & -\frac{1}{2} & & & \ddots & \\ & & & \frac{1}{2} & & 1 \end{array}\right) & \begin{array}{c} 1 \\ 2 \\ \vdots \\ m \\ m+1 \\ \vdots \\ k \end{array} \end{array}$$

and the integer m is $k/2$ or $(k+1)/2$.

Table 2

Equicoordinate percentage points h of a multivariate t distribution with correlation matrix Σ

for

$P = 0.50, 0.75, 0.90, 0.95, 0.975, 0.99;$

$k = 2(1)6(2)12(4)20;$ and

$\text{D.F.} = \nu = 5(1)10(2)20(4)60(30)120.$

Note: The table value $h = h_\nu(P, k)$ is the solution of the equation

$$P = \frac{1}{(2\pi)^{k/2}|\Sigma|^{\frac{1}{2}}} \int_{-\infty}^{h}\int_{-\infty}^{h}\cdots\int_{-\infty}^{h}\left[\int_{0}^{\infty} e^{-\frac{u^2}{2}\,\underset{\sim}{t}'\Sigma^{-1}\underset{\sim}{t}} \frac{2(\frac{\nu}{2})^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} u^{k+\nu-1} e^{-\frac{\nu}{2}u^2} du\right] \prod_{i=1}^{k} dt_i$$

or, equivalently,

$$P = \frac{\Gamma(\frac{k+\nu}{2})}{(\nu\pi)^{\frac{k}{2}}|\Sigma|^{\frac{1}{2}}\Gamma(\frac{\nu}{2})} \int_{-\infty}^{h}\int_{-\infty}^{h}\cdots\int_{-\infty}^{h} \left[1 + \frac{1}{\nu}\,\underset{\sim}{t}'\Sigma^{-1}\underset{\sim}{t}\right]^{-\frac{(k+\nu)}{2}} \prod_{i=1}^{k} dt_i$$

where Σ is defined in the note of Table 1.

Table 2A

P = 0.50

D.F.	K = 2	K = 3	K = 4	K = 5	K = 6	K = 8	K = 10	K = 12	K = 16	K = 20
5	.68057	.89079	1.06270	1.16737	1.26195	1.39388	1.49125	1.56779	1.68341	1.76905
6	.67397	.88145	1.05075	1.15380	1.24685	1.37666	1.47249	1.54784	1.66173	1.74613
7	.66931	.87488	1.04234	1.14425	1.23620	1.36449	1.45921	1.53371	1.64633	1.72982
8	.66584	.87000	1.03610	1.13716	1.22829	1.35544	1.44931	1.52315	1.63480	1.71760
9	.66317	.86623	1.03129	1.13168	1.22218	1.34843	1.44163	1.51495	1.62583	1.70807
10	.66104	.86324	1.02746	1.12733	1.21731	1.34284	1.43551	1.50840	1.61865	1.70043
12	.65786	.85878	1.02176	1.12084	1.21005	1.33449	1.42633	1.49857	1.60784	1.68892
14	.65561	.85562	1.01771	1.11623	1.20489	1.32854	1.41978	1.49154	1.60009	1.68064
16	.65393	.85326	1.01470	1.11279	1.20103	1.32408	1.41486	1.48626	1.59425	1.67438
18	.65263	.85143	1.01236	1.11012	1.19804	1.32061	1.41104	1.48214	1.58969	1.66949
20	.65159	.84998	1.01049	1.10799	1.19565	1.31784	1.40797	1.47884	1.58602	1.66555
24	.65003	.84780	1.00770	1.10480	1.19207	1.31369	1.40337	1.47388	1.58050	1.65961
28	.64892	.84625	1.00571	1.10253	1.18952	1.31072	1.40008	1.47032	1.57653	1.65532
32	.64810	.84509	1.00423	1.10083	1.18760	1.30850	1.39761	1.46765	1.57354	1.65209
36	.64745	.84419	1.00307	1.09951	1.18612	1.30677	1.39568	1.46557	1.57120	1.64956
40	.64694	.84347	1.00215	1.09846	1.18493	1.30538	1.39414	1.46389	1.56933	1.64753
44	.64652	.84288	1.00140	1.09760	1.18396	1.30424	1.39288	1.46253	1.56779	1.64586
48	.64617	.84239	1.00077	1.09688	1.18315	1.30330	1.39182	1.46138	1.56651	1.64447
52	.64588	.84198	1.00024	1.09627	1.18246	1.30250	1.39093	1.46041	1.56542	1.64328
56	.64562	.84163	0.99979	1.09575	1.18188	1.30181	1.39017	1.45958	1.56448	1.64227
60	.64540	.84132	0.99939	1.09530	1.18137	1.30122	1.38950	1.45886	1.56367	1.64138
90	.64438	.83989	0.99756	1.09320	1.17900	1.29844	1.38640	1.45549	1.55986	1.63723
120	.64387	.83918	0.99664	1.09215	1.17781	1.29705	1.38485	1.45380	1.55795	1.63515
∞	.64235	.83705	0.99390	1.08901	1.17426	1.29288	1.38017	1.44870	1.55216	1.62881

Table 2B

p = 0.75

D.F.	K = 2	K = 3	K = 4	K = 5	K = 6	K = 8	K = 10	K = 12	K = 16	K = 20
5	1.28547	1.49920	1.67149	1.78273	1.88216	2.02501	2.13209	2.21722	2.34734	2.44484
6	1.26035	1.46657	1.63178	1.73848	1.83356	1.97019	2.07255	2.15393	2.27832	2.37156
7	1.24291	1.44395	1.60427	1.70782	1.79988	1.93216	2.03122	2.10994	2.23027	2.32047
8	1.23009	1.42735	1.58409	1.68534	1.77517	1.90423	2.00083	2.07757	2.19487	2.28279
9	1.22027	1.41465	1.56866	1.66814	1.75626	1.88285	1.97754	2.05276	2.16769	2.25384
10	1.21251	1.40463	1.55647	1.65455	1.74132	1.86594	1.95913	2.03312	2.14617	2.23088
12	1.20102	1.38980	1.53845	1.63447	1.71923	1.84093	1.93185	2.00402	2.11422	2.19678
14	1.19293	1.37936	1.52577	1.62033	1.70367	1.82330	1.91262	1.98348	2.09165	2.17266
16	1.18693	1.37161	1.51636	1.60984	1.69212	1.81020	1.89832	1.96821	2.07484	2.15469
18	1.18229	1.36564	1.50910	1.60174	1.68321	1.80010	1.88728	1.95640	2.06185	2.14078
20	1.17860	1.36089	1.50333	1.59531	1.67612	1.79205	1.87850	1.94701	2.05150	2.12969
24	1.17311	1.35381	1.49473	1.58572	1.66556	1.78007	1.86540	1.93300	2.03605	2.11313
28	1.16921	1.34879	1.48864	1.57892	1.65807	1.77157	1.85610	1.92305	2.02507	2.10136
32	1.16631	1.34505	1.48409	1.57385	1.65248	1.76522	1.84915	1.91561	2.01687	2.09256
36	1.16405	1.34215	1.48056	1.56992	1.64815	1.76030	1.84377	1.90985	2.01050	2.08573
40	1.16226	1.33983	1.47775	1.56678	1.64470	1.75637	1.83948	1.90525	2.00543	2.08028
44	1.16079	1.33795	1.47546	1.56422	1.64188	1.75317	1.83597	1.90150	2.00128	2.07583
48	1.15957	1.33638	1.47355	1.56210	1.63954	1.75051	1.83305	1.89838	1.99783	2.07212
52	1.15854	1.33505	1.47194	1.56030	1.63755	1.74825	1.83059	1.89574	1.99491	2.06899
56	1.15766	1.33392	1.47056	1.55876	1.63586	1.74633	1.82848	1.89348	1.99241	2.06631
60	1.15689	1.33293	1.46937	1.55743	1.63439	1.74466	1.82666	1.89152	1.99025	2.06399
90	1.15334	1.32837	1.46382	1.55124	1.62757	1.73690	1.81816	1.88242	1.98019	2.05318
120	1.15158	1.32609	1.46106	1.54815	1.62417	1.73304	1.81393	1.87789	1.97518	2.04779
∞	1.14630	1.31930	1.45281	1.53895	1.61402	1.72150	1.80130	1.86435	1.96020	2.03170

Table 2C

P = 0.90

D.F.	K = 2	K = 3	K = 4	K = 5	K = 6	K = 8	K = 10	K = 12	K = 16	K = 20
5	2.00771	2.24189	2.42873	2.55442	2.66588	2.82937	2.95328	3.05255	3.20554	3.32111
6	1.93795	2.15646	2.32910	2.44550	2.54818	2.69895	2.81313	2.90461	3.04560	3.15216
7	1.89059	2.09861	2.26174	2.37189	2.46867	2.61082	2.71840	2.80455	2.93732	3.03766
8	1.85635	2.05686	2.21320	2.31886	2.41140	2.54735	2.65015	2.73244	2.85922	2.95502
9	1.83046	2.02534	2.17657	2.27887	2.36821	2.49948	2.59867	2.67803	2.80026	2.89261
10	1.81019	2.00070	2.14796	2.24764	2.33449	2.46210	2.55846	2.63554	2.75419	2.84381
12	1.78053	1.96468	2.10617	2.20203	2.28526	2.40754	2.49977	2.57348	2.68688	2.77248
14	1.75987	1.93963	2.07713	2.17035	2.25106	2.36964	2.45899	2.53036	2.64009	2.72288
16	1.74465	1.92120	2.05578	2.14707	2.22594	2.34180	2.42904	2.49868	2.60570	2.68640
18	1.73298	1.90707	2.03943	2.12924	2.20670	2.32048	2.40610	2.47442	2.57936	2.65846
20	1.72375	1.89590	2.02650	2.11514	2.19149	2.30364	2.38798	2.45525	2.55854	2.63637
24	1.71006	1.87936	2.00737	2.09429	2.16900	2.27872	2.36117	2.42689	2.52775	2.60369
28	1.70041	1.86770	1.99389	2.07961	2.15316	2.26118	2.34229	2.40693	2.50606	2.58068
32	1.69324	1.85904	1.98388	2.06870	2.14141	2.24816	2.32828	2.39211	2.48997	2.56360
36	1.68770	1.85236	1.97616	2.06029	2.13233	2.23811	2.31748	2.38068	2.47756	2.55042
40	1.68329	1.84704	1.97002	2.05360	2.12512	2.23013	2.30889	2.37160	2.46769	2.53995
44	1.67970	1.84271	1.96502	2.04815	2.11925	2.22363	2.30190	2.36420	2.45966	2.53142
48	1.67672	1.83911	1.96087	2.04364	2.11438	2.21823	2.29610	2.35807	2.45299	2.52435
52	1.67420	1.83608	1.95737	2.03982	2.11027	2.21369	2.29121	2.35289	2.44737	2.51838
56	1.67205	1.83349	1.95438	2.03657	2.10676	2.20980	2.28703	2.34847	2.44257	2.51329
60	1.67019	1.83125	1.95179	2.03375	2.10373	2.20644	2.28342	2.34465	2.43842	2.50888
90	1.66157	1.82086	1.93980	2.02070	2.08967	2.19088	2.26668	2.32695	2.41920	2.48847
120	1.65729	1.81571	1.93386	2.01423	2.08270	2.18316	2.25838	2.31818	2.40967	2.47836
∞	1.64457	1.80040	1.91622	1.99504	2.06202	2.16029	2.23378	2.29217	2.38142	2.44838

Table 2D

P = 0.95

D.F.	K = 2	K = 3	K = 4	K = 5	K = 6	K = 8	K = 10	K = 12	K = 16	K = 20
5	2.56513	2.82213	3.02640	3.16640	3.29014	3.47342	3.61302	3.72530	3.89898	4.03064
6	2.44344	2.67692	2.86038	2.98658	3.09743	3.26186	3.38705	3.48772	3.64352	3.76174
7	2.36222	2.58029	2.75012	2.86724	2.96961	3.12161	3.23724	3.33021	3.47406	3.58324
8	2.30424	2.51145	2.67169	2.78241	2.87879	3.02198	3.13082	3.21829	3.35361	3.45629
9	2.26079	2.45996	2.61310	2.71907	2.81101	2.94764	3.05142	3.13479	3.26370	3.36150
10	2.22705	2.42002	2.56770	2.67001	2.75851	2.89008	2.98995	3.07013	3.19408	3.28809
12	2.17806	2.36212	2.50195	2.59900	2.68257	2.80684	2.90105	2.97663	3.09338	3.18186
14	2.14423	2.32220	2.45668	2.55011	2.63031	2.74958	2.83991	2.91232	3.02410	3.10878
16	2.11947	2.29301	2.42361	2.51443	2.59218	2.70781	2.79531	2.86541	2.97357	3.05546
18	2.10056	2.27075	2.39840	2.48723	2.56313	2.67600	2.76134	2.82970	2.93509	3.01485
20	2.08566	2.25322	2.37856	2.46583	2.54027	2.65097	2.73463	2.80160	2.90483	2.98291
24	2.06367	2.22736	2.34932	2.43430	2.50660	2.61412	2.69529	2.76024	2.86027	2.93589
28	2.04822	2.20921	2.32881	2.41220	2.48300	2.58829	2.66774	2.73126	2.82906	2.90295
32	2.03678	2.19577	2.31363	2.39584	2.46554	2.56919	2.64736	2.70984	2.80598	2.87860
36	2.02796	2.18542	2.30195	2.38325	2.45211	2.55450	2.63168	2.69335	2.78823	2.85986
40	2.02095	2.17721	2.29267	2.37326	2.44145	2.54284	2.61924	2.68028	2.77415	2.84500
44	2.01526	2.17053	2.28514	2.36514	2.43278	2.53337	2.60914	2.66966	2.76271	2.83293
48	2.01053	2.16499	2.27889	2.35841	2.42560	2.52552	2.60077	2.66085	2.75323	2.82292
52	2.00655	2.16032	2.27362	2.35274	2.41956	2.51891	2.59371	2.65344	2.74525	2.81450
56	2.00315	2.15634	2.26913	2.34790	2.41439	2.51326	2.58769	2.64711	2.73843	2.80731
60	2.00021	2.15289	2.26525	2.34372	2.40993	2.50839	2.58250	2.64165	2.73255	2.80110
90	1.98661	2.13696	2.24729	2.32438	2.38931	2.48585	2.55846	2.61638	2.70534	2.77239
120	1.97987	2.12908	2.23840	2.31482	2.37912	2.47470	2.54658	2.60390	2.69190	2.75820
∞	1.95993	2.10574	2.21213	2.28654	2.34897	2.44177	2.51147	2.56700	2.65218	2.71629

Table 2E

P = 0.975

D.F.	K = 2	K = 3	K = 4	K = 5	K = 6	K = 8	K = 10	K = 12	K = 16	K = 20
5	3.15876	3.44437	3.67090	3.82816	3.96687	4.17376	4.33198	4.45962	4.65777	4.80857
6	2.96601	3.21856	3.41633	3.55432	3.67516	3.85575	3.99376	4.10505	4.27776	4.40917
7	2.83955	3.07086	3.25020	3.37579	3.48516	3.64884	3.77385	3.87465	4.03111	4.15022
8	2.75037	2.96695	3.13355	3.25053	3.35193	3.50382	3.61976	3.71320	3.85821	3.96860
9	2.68419	2.88996	3.04725	3.15792	3.25348	3.39672	3.50597	3.59398	3.73053	3.83445
10	2.63315	2.83069	2.98088	3.08673	3.17784	3.31446	3.41859	3.50245	3.63250	3.73145
12	2.55965	2.74547	2.88560	2.98458	3.06935	3.19655	3.29337	3.37128	3.49201	3.58382
14	2.50931	2.68721	2.82053	2.91488	2.99537	3.11618	3.20804	3.28191	3.39631	3.48325
16	2.47269	2.64487	2.77332	2.86432	2.94172	3.05793	3.14622	3.21717	3.32699	3.41040
18	2.44487	2.61273	2.73750	2.82598	2.90106	3.01379	3.09938	3.16813	3.27448	3.35522
20	2.42301	2.58751	2.70941	2.79591	2.86918	2.97921	3.06268	3.12971	3.23335	3.31200
24	2.39088	2.55046	2.66818	2.75181	2.82244	2.92850	3.00889	3.07340	3.17307	3.24866
28	2.36840	2.52457	2.63939	2.72102	2.78981	2.89312	2.97137	3.03413	3.13104	3.20450
32	2.35179	2.50545	2.61814	2.69831	2.76575	2.86704	2.94372	3.00519	3.10007	3.17196
36	2.33903	2.49076	2.60183	2.68087	2.74728	2.84702	2.92249	2.98297	3.07630	3.14699
40	2.32891	2.47912	2.58890	2.66705	2.73265	2.83117	2.90569	2.96539	3.05749	3.12722
44	2.32069	2.46967	2.57841	2.65584	2.72078	2.81831	2.89206	2.95112	3.04223	3.11119
48	2.31388	2.46184	2.56972	2.64656	2.71096	2.80766	2.88077	2.93932	3.02960	3.09792
52	2.30815	2.45525	2.56241	2.63875	2.70269	2.79871	2.87128	2.92939	3.01897	3.08676
56	2.30326	2.44963	2.55618	2.63209	2.69564	2.79107	2.86319	2.92092	3.00991	3.07724
60	2.29903	2.44478	2.55079	2.62634	2.68955	2.78448	2.85620	2.91361	3.00209	3.06903
90	2.27951	2.42236	2.52594	2.59979	2.66146	2.75406	2.82396	2.87988	2.96602	3.03114
120	2.26987	2.41129	2.51367	2.58669	2.64760	2.73905	2.80807	2.86326	2.94824	3.01246
∞	2.24140	2.37864	2.47751	2.54808	2.60676	2.69486	2.76125	2.81429	2.89587	2.95746

Table 2F

P = 0.99

D.F.	K = 2	K = 3	K = 4	K = 5	K = 6	K = 8	K = 10	K = 12	K = 16	K = 20
5	4.02790	4.36051	4.62385	4.80875	4.97154	5.21579	5.40324	5.55485	5.79095	5.97127
6	3.70523	3.98906	4.21075	4.36746	4.50437	4.71040	4.86846	4.99628	5.19528	5.34717
7	3.49824	3.75150	3.94718	4.08617	4.20682	4.38875	4.52823	4.64100	4.81657	4.95059
8	3.35462	3.58711	3.76519	3.89213	4.00177	4.16733	4.29419	4.39675	4.55641	4.67835
9	3.24933	3.46683	3.63226	3.75052	3.85222	4.00594	4.12366	4.21879	4.36687	4.47994
10	3.16892	3.37513	3.53105	3.64276	3.73848	3.88327	3.99408	4.08360	4.22289	4.32923
12	3.05435	3.24470	3.38732	3.48983	3.57717	3.70941	3.81051	3.89211	4.01901	4.11584
14	2.97673	3.15650	3.29029	3.38666	3.46843	3.59230	3.68690	3.76322	3.88182	3.97227
16	2.92071	3.09293	3.22044	3.31244	3.39023	3.50814	3.59811	3.67065	3.78331	3.86919
18	2.87839	3.04496	3.16779	3.25651	3.33134	3.44478	3.53128	3.60099	3.70920	3.79165
20	2.84531	3.00748	3.12669	3.21287	3.28540	3.39538	3.47919	3.54669	3.65144	3.73122
24	2.79692	2.95274	3.06669	3.14918	3.21839	3.32335	3.40325	3.46756	3.56729	3.64320
28	2.76325	2.91467	3.02501	3.10497	3.17188	3.27338	3.35058	3.41270	3.50896	3.58219
32	2.73848	2.88669	2.99439	3.07248	3.13773	3.23669	3.31193	3.37243	3.46616	3.53742
36	2.71948	2.86524	2.97093	3.04761	3.11158	3.20862	3.28235	3.34162	3.43341	3.50318
40	2.70446	2.84829	2.95240	3.02796	3.09093	3.18644	3.25899	3.31730	3.40756	3.47614
44	2.69228	2.83455	2.93738	3.01204	3.07420	3.16848	3.24008	3.29760	3.38663	3.45425
48	2.68221	2.82318	2.92497	2.99888	3.06037	3.15364	3.22445	3.28133	3.36933	3.43617
52	2.67374	2.81363	2.91453	2.98783	3.04875	3.14118	3.21132	3.26766	3.35481	3.42099
56	2.66651	2.80549	2.90564	2.97841	3.03886	3.13056	3.20014	3.25601	3.34244	3.40805
60	2.66029	2.79847	2.89798	2.97028	3.03032	3.12140	3.19049	3.24597	3.33177	3.39690
90	2.63157	2.76612	2.86267	2.93287	2.99103	3.07925	3.14612	3.19977	3.28269	3.34559
120	2.61743	2.75019	2.84529	2.91447	2.97171	3.05853	3.12430	3.17706	3.25857	3.32038
∞	2.57583	2.70340	2.79428	2.86045	2.91500	2.99774	3.06033	3.11048	3.18787	3.24648

BIBLIOGRAPHY

[1] Anscombe, F. J. (1952). Large sample theory of sequential estimation. Proc. Cambridge Philos. Soc., 48, 600-607.

[2] Anscombe, F. J. (1953). Sequential estimation. J. Roy. Stat. Soc., Ser. B, 15, 1-21.

[3] Bechhofer, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. Ann. Math. Statist., 25, 16-39.

[4] Bechhofer, R. E., Dunnett, C. W. and Sobel, M. (1954). A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance. Biometrika, 41, 170-176.

[5] Bechhofer, R. E., Kiefer, J. and Sobel, M. (1967). Sequential Indentification and Ranking Procedures. Statist. Res. Monog. Ser., Inst. Math. Statist. and Univ. Chicago Press, Chicago.

[6] Bechhofer, R. E. and Sobel, M. (1954). A single-sample multiple decision procedure for ranking variances of normal populations. Ann. Math. Statist., 25, 273-289.

[7] Chow, Y. S. and Robbins, H. (1965). On the asymptotic theory of fixed-width sequential confidence intervals for the mean. Ann. Math. Statist., 36, 457-462.

[8] Cramér, H. (1946). Mathematical Methods of Statistics. Princeton University Press, New Jersey.

[9] Dantzig, G. B. (1940). On the non-existence of tests of "Student's" hypothesis having power functions independent of σ. Ann. Math. Statist., 11, 186-190.

[10] Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. J. Amer. Stat. Assn., 50, 1096-1121.

[11] Dunnett, C. W. and Sobel, M. (1954). A bivariate generalization on Student's t-distribution, with tables for certain special cases. Biometrika, 41, 153-169.

[12] Eaton, M. L. (1966). Some optimal properties of ranking procedures with applications in multivariate analysis. Technical Report No. 22, Department of Statistics, Stanford University.

[13] Feller, W. (1961). An Introduction to Probability Theory and its Applications, Vol. I (2nd. Ed.). John Wiley and Sons, Inc., New York.

[14] Gleser, L. J.. Robbins, H. and Starr, N. (1964). Some asymptotic properties of fixed-width sequential confidence intervals for the mean of a normal population with unknown variance. Columbia University Report.

[15] Gupta, S. S. (1963). Probability integrals of multivariate normal and multivariate t. Ann. Math. Statist., 34, 792-828.

[16] Gupta, S. S. and Sobel, M. (1957). On a statistic which arises in selection and ranking problems. Ann. Math. Statist., 28, 957-967.

[17] Gupta, S. S. and Sobel, M. (1958). On selecting a subset which contains all populations better than a standard. Ann. Math. Statist., 29, 235-244.

[18] Lehmann, E. L. (1957). A theory of some multiple decision problems, I. Ann. Math. Statist., 28, 1-25.

[19] Lehmann, E. L. (1961). Some model I problems of selection. Ann. Math. Statist., 32, 990-1012.

[20] Lehmann, E. L. (1962). Notes on the Theory of Estimation. University of California, Berkeley, California.

[21] Loève, M. (1963). Probability Theory (3rd. Ed.). D. Van Nostrand Co., New Jersey.

[22] Milton, R. C. (1963). Tables of the equally correlated multivariate normal probability integral. Technical Report No. 27, Department of Statistics, University of Minnesota.

[23] Paulson, E. (1952). On the comparison of several experimental categories with a control. Ann. Math. Statist., 23, 239-246.

[24] Ray, W. D. (1957). Sequential confidence intervals for the mean of a normal population with unknown variance. J. Roy. Stat. Soc., Ser. B, 19, 133-143.

[25] Robbins, H. (1959). Sequential estimation of the mean of a normal population. Probability and Statistics (Harald Cramér Volume). Almquist and Wiksell, Uppsala, Sweden, 235-245.

[26] Robbins, H., Sobel, M. and Starr, N. (1967). A sequential procedure for selecting the largest of k means. Joint Technical Report No. 88, Department of Statistics, University of Minnesota; No. 4, Department of Statistics, Carnegie Institute of Technology.

[27] Robbins, H. and Starr, N. (1965). Remarks on sequential hypothesis testing. Technical Report No. 68, Department of Statistics, University of Minnesota.

[28] Seelbinder, B. M. (1953). On Stein's two-stage sampling scheme. Ann. Math. Statist., 24, 640-649.

[29] Starr, N. (1966). The performance of a sequential procedure for the fixed-width interval estimation of the mean. Ann. Math. Statist., 37, 36-50.

[30] Stein, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. Ann. Math. Statist., 16, 243-258.

[31] Stroud, A. H. and Secrest, D. (1966). Gaussian Quadrature Formulas. Prentice-Hall, Inc., New Jersey.